Extension to lower energies of the cosmic-ray energy window at the Pierre Auger Observatory

Messina, Stefano

Extension to lower energies of the cosmic-ray energy window at the Pierre Auger Observatory

Stefano Antonino Maria Messina

The cover image is adapted from an aereal picture of the artwork "Anni-luce" (Giulio Paolini, 88 × 136.5 m, 2000-01, granites and stones of different origin, diorite of Anzola, Carrara marble), realized in the yard of the Civic Gallery of Modern and Contemporary Art (GAM) in Turin, Italy (after courtesy of the Giulio a Anna Paolini Foundation). On the ground, that opening towards the sky, with papers fluttering as if *attracted by the infinite*, elements of the cosmo mixed with earthly elements, has a powerful connection to what researchers do: remove the roof from our mind, often so closed, to investigate the mysteries of the cosmos, the immensity of the Universe within ourselves.

university of groningen

FOM

NWO
Nederlandse Organisatie voor Wetenschappelijk Onderzoek

This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO). The printing has been financially supported by the Rijksuniversiteit Groningen.

COVER DESIGN: Raffaello Buccheri
PRINTED BY: IPSKAMPPRINTING.NL
ISBN: 978-94-028-0286-3

university of groningen

Extension to lower energies of the cosmic-ray energy window at the Pierre Auger Observatory

PhD Thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 23 September 2016 at 11.00 hours

by

Stefano Messina

born on 28 August 1984
in Menfi, Italy

Supervisors
Prof. A. M. van den Berg
Prof. O. Scholten

Assessment Committee
Prof. N. Kalantar-Nayestanaki
Prof. D. Z. Besson
Prof. S. Zaroubi

Dedicated to Mariasole and Margherita, my life would have not been life without you.

To my Family, who has unconditionally supported and still supports me.

To my Friends, with whom I've shared this wonderful part of existence.

"When I feel about myself, I feel there's the unknown, there're things I don't know and that's what I'm striving for…

So, to say that I know, will not be true because there're things I don't know and I'm trying to find out what they are…And they come little by little…"

Yusef Lateef, Brother Yusef (doc.)

Contents

1 Introduction **1**
1.1 A Glimpse of the Cosmos 1
1.2 Historical Overview 4
1.3 Thesis Outlook 6

2 Cosmic-Ray Phenomenology **9**
2.1 Ultra-High-Energy Cosmic Rays 9
2.1.1 Spectral Features 10
The Dip Region 11
The GZK cutoff 13
2.1.2 Origin of UHECR 14
Acceleration Mechanisms 14
Sources 15
Propagation 17
2.2 Extensive Air Showers 17
2.2.1 Dynamics of Air-Shower Development 20
Longitudinal Profile 20
Lateral Distribution 25
Radio Emission 28
2.3 Final Remarks 32

3 The Pierre Auger Observatory **35**
3.1 A Hybrid Conception 36
3.2 Fluorescence Detector 36
3.2.1 The HEAT enhancement 39
3.3 Surface Detector 39
Detector Calibration 40
Surface Detector Trigger 42

3.3.1 The AMIGA enhancement 44
3.4 Radio Detector: AERA 45
3.4.1 The AERAlet enhancement 47
3.5 Conclusion 48

4 Reconstruction of Cosmic-Ray Events 51
4.1 The Offline Framework 51
4.2 SD-Only Reconstruction of Cosmic-Ray events 52
4.2.1 Reconstruction algorithms 54
Planar approximation 54
Determining the LDF 56
4.2.2 Data Storage and Visualisation 58
4.3 Radio-SD Reconstruction 59
4.3.1 Polarization Information 61
4.4 Summary 63

5 AERAlet Measurements of Cosmic Rays 65
5.1 AERAlet Monitoring and Performances 66
5.2 Tuning the LDF Parameters 67
5.2.1 Event Selection 68
5.2.2 Event Reconstruction 69
5.2.3 Data Analysis 70
Optimal Distance 70
Slope Parametrisation 73
5.2.4 Uncertainty and Bias Analysis 76
5.3 Energy Calibration 83
5.3.1 Energy Estimator for AERAlet 83
Energy-Estimator Uncertainty 90
5.3.2 Event Selection and Reconstruction 91
Geometrical-Reconstruction Comparison 92
5.3.3 Toy Monte Carlo 94
Implementation 96
5.3.4 Estimation of the Calibration Parameters 99
5.4 Low-Energy Extension of the Energy Spectrum 102
5.4.1 Exposure Calculation 105
5.4.2 Heuristic Correction 108

5.4.3 AERAlet Energy Spectrum 111
5.5 Conclusions 117

6 Measurement Sensitivity of the Radially-Polarized Radio Emission Component 119
6.1 Event Selection and Reconstruction 119
6.1.1 Antenna-wise charge-excess fraction 120
6.1.2 Observer-Angle Uncertainty 123
6.1.3 Event and Station selection 126
6.2 Modelling the amount of Radial Polarization 129
6.2.1 Bootstrap Approach 135
6.3 Conclusions 142

7 Summary and Discussion 145

A Configuration files for $\overline{\text{Off}}\underline{\text{line}}$ modules 149
A.1 SD-Only Reconstruction Pipeline 149
A.2 Radio-SD Reconstruction Pipeline 150
A.3 `SdEventSelectorOG` module 151
A.4 `LDFFinderKG` module 152

B Estimation of the Variance for angular variables 157
B.1 Variance Estimation 157

List of Publications 159

Acronyms 165

List of Tables 167

List of Figures 168

Bibliography 171

Samenvatting 181

"First of all was the Chaos . . . "

Hesiod, Theogony

"Whence things have their origin, thence also their destruction happens, according to necessity; For they give to each other justice and recompense for their injustice in conformity with the ordinance of Time."

Anaximander, Fragments

1 Introduction

This chapter is mainly intended to offer an overview of the motivations that drive scientists all around the world to spend efforts in the Astroparticle Physics field of research, and in particular unrevealing the mysteries of Ultra-High Energy Cosmic Rays (UHECRs).

1.1 A Glimpse of the Cosmos

The word *cosmos* comes from the Greek κόσμος that originally meant "order", and only successively was identified as the ordered Universe originated from the "formless mass", the χάος. The origin of the Universe is probably one of the first questions that has popped up in humans' mind, once they started to observe the sky with the naked eye. As a matter of fact, in *comprehending* the Universe, its origin and all related processes, astronomical observations have played a key role from the very beginning. Among all different observation techniques, the detection of *cosmic diffuse radiations* is an exceptional scientific experience that offers us quite a comprehensive picture of the Universe beyond the window of visible light.

Here, 'cosmic diffuse radiations' refers in a broad sense to the various radiation fields and particles filling up our Universe. As a matter of fact, all

space is permeated by a wide variety of diffuse electromagnetic and particle radiations sharing the feature of being *cosmic*. Such radiations and particles offer us a "glimpse" of what are and have been the astrophysical processes driving the evolution of Universe, and structure formation processes within it. In other words, the nature and history of the Universe are "coded" in these radiation fields and particles, so any attempt of understanding Universe origin and evolution processes must be able to describe them.

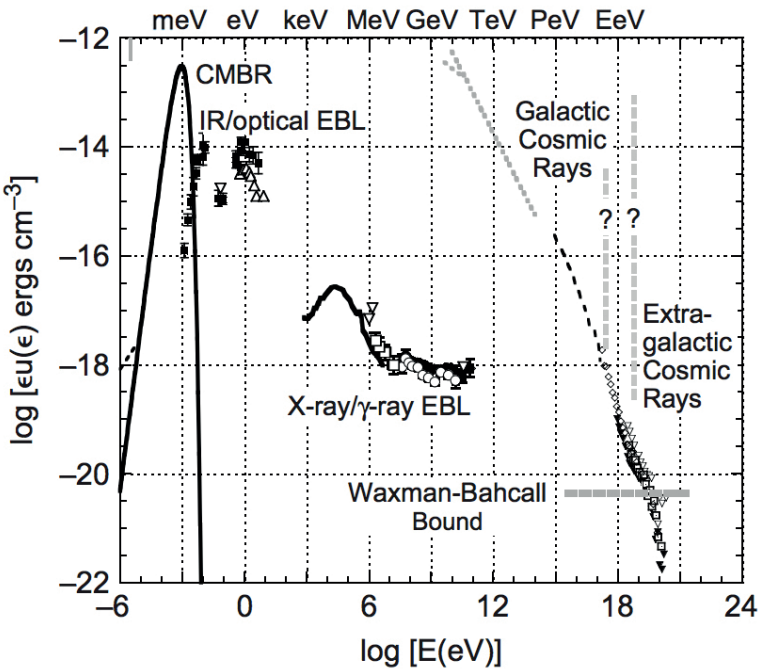

FIGURE 1.1: Energy-density spectrum of the different kinds of diffuse background radiations and particles permeating the Universe. Image taken from Ref. [1]. The Waxman-Bahcall bound provides an upper limit on the high-energy neutrino flux produced by photo-meson interactions of UHECRs with the CMBR (see e.g. [2]).

Fig. 1.1 shows the energy-density spectrum of the diverse contributions to this diffuse glow. At the far left of the spectrum, i.e. at the lowest energies, radio background and the Cosmic Microwave Background Radiation (CMBR) ($\sim 2.7^{\circ}$ K) dominate. The different contributions to the Extragalactic Background Light (EBL), such as Infra-Red (IR) background, optical light, γ-rays, and X-rays, are situated in the central part of the spectrum, whilst at the far right, i.e. at the highest energies, radiation made up by particles, i.e. Cosmic Rays (CRs) and cosmogenic neutrino's,

are the only diffuse radiations of which we do not know the sources.

Each of these contributions necessarily reflects a peculiar process. The CMBR has been related to the earliest processes of structure formation, whilst infrared light is mainly due to stellar light backscattered by the cold dust of the interstellar medium. Optical light is primarily due to stellar radiation, and contributions to the X-ray radiation come from star-formation and accretion-disk regions. Finally, very-high energy γ-rays, CRs and neutrino's carry highly valuable information about non-thermal processes taking place "somewhere" in the Universe.

Indeed, the existence/presence of such high-energy particles, entering Earth's atmosphere from outer space, bear witnesses of the importance of non-thermal phenomena in the Universe. In particular, the origin of the highest energy particles in the Universe (at the $\sim$ EeV[1] regime and beyond), known as UHECRs, is one of the prominent challenges in modern astrophysics. Indeed, UHECR arrival directions (cf. Fig. 1.2) are consistent with the isotropic hypothesis, and though the astrophysical sources capable of accelerating particles to extreme energies are assumed to be few, no obvious correlation with the local luminous mass distribution has been found yet [3]. In addition, a transition from galactic to extragalactic cosmic rays is assumed to take place at the *dip* (cf. the vertical dashed lines in Fig. 1.1), due to the limited confinement efficiency of the galactic magnetic field. At which energy this transition occurs, and its dependency on CR composition is recognized as a key issue in UHECR research.

The dip region, indicated by the vertical-dashed grey lines in Fig. 1.1, extends approximately from 10^{17} to 10^{19} eV. In this range of energies, one observes several variations (steepening or flattening) from the average spectrum slope, known as spectral features, either correlated with the transition from galactic to extragalactic sources or with a change of the chemical composition (or most likely with both). As a matter of fact, some of these features occur around the lower limit of 10^{17} eV, and are poorly studied [4]. However the current energy threshold of the Pierre Auger Observatory is at $3 \cdot 10^{17}$ eV. To have a clear picture of the energy range where the aforementioned transition is assumed to occur, it is necessary to lower this energy threshold. The extension of the cosmic-ray energy window at the Observatory is of high priority, and the present study set the basis for such a goal.

[1] 1 EeV $= 10^{18}$ eV $= 1.602176565 \cdot 10^{-1}$ Joule

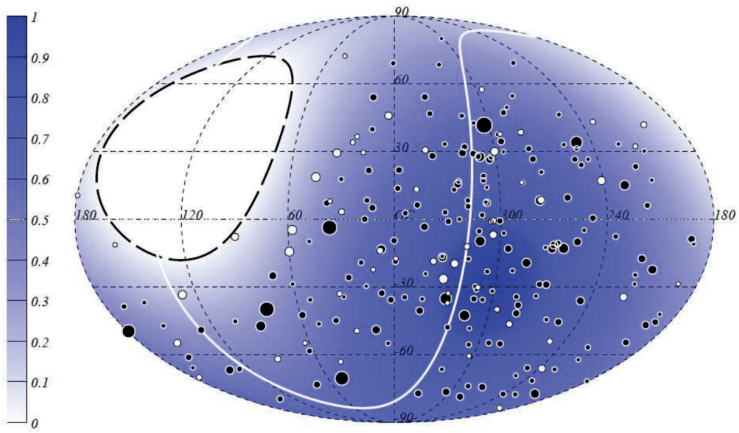

FIGURE 1.2: Map in galactic coordinates of the arrival directions of UHECRs with primary energy $\geqslant 52{\cdot}10^{18}$ eV detected at the Pierre Auger Observatory in Argentina. The black (white) circles correspond to "vertical" ("inclined") events, i.e. arriving with a zenith angle $\theta \leqslant 60^{\circ} (> 60^{\circ})$. The size of each circle scales with the energy of the event. The colour scale is proportional to the relative exposure. The black-dashed line is the field-of-view limit for the Auger Observatory (for $\theta \leqslant 80^{\circ}$) and the white-solid line corresponds to the Super-Galactic Plane. Image taken from Ref. [3].

1.2 Historical Overview

Since their discovery by Victor Hess in 1912 [5], CRs have never ceased to arouse interest. The mysteries related to their existence always led to significant experimental efforts in building new detectors. Many of the great discoveries in particle physics came from the observation of cosmic rays. Indeed, positrons, muons and pions, were discovered measuring this "radiation coming from above" [6, 7, 8].

In 1938 Pierre Auger made his great discovery: positioning particle counters at roughly the same height above sea level with a certain distance between them, these counters sometimes were simultaneously triggered. This experiment demonstrated that some particles were arriving in groups at the surface of the Earth, and that the observed particles were secondaries from a common source. Pierre Auger was the first to conclude that particles, detected in time coincidence, were produced from a single (*primary*) particle, interacting with the air molecules and generating many and different particles. Such cascades, with billions of particles, are distributed over areas of 10-100 square kilometres, hence the name *Extensive Air Showers (EASs)* [9].

Soon, he realized that particles exist in Nature with very high energies,

above 10^{15} eV and beyond. After this discovery, large arrays of detectors were developed to study extensive air showers. In the sixties, the largest detector array had been installed by the MIT group at Volcano Ranch (New Mexico). It was covering an area of 12 km^2 with 20 stations in a triangular grid, and it showed the existence of *UHECRs* by observing an incredibly large air shower generated by a $\sim 10^{20}$ eV particle [10].

Hitherto, since the mid 1940s, counters were primarily made exploiting scintillator materials. Alternative instruments were constructed using a medium, such as water, in which particles generate light through the Cherenkov effect. Such detectors, were developed in the fifties and some years after, physicists have turned also to the possibility of observing the light generated by the fast secondary particles in the atmosphere itself. Atmospheric fluorescence light, was observed by W. Galbraith and J.V. Jelley in 1953 [11], and subsequent studies brought important advances in cosmic-ray research. As this fluorescence light signal is very weak, it can only be observed during dark nights, limiting the uptime for this detection technique. Then, to avoid this problem, Jelley himself started to consider the option to change the frequency of observation from the optical to the radio band. The conclusion he reached by considering this option was not encouraging at all, but the theoretical results of G.A. Askaryan in 1962/1963 led him to make a great discovery in 1964: the first observation of radio pulses from extensive air showers [12]. Despite the initial success, the radio-detection technique of cosmic rays was abandoned because of technical difficulties and because interest moved to cm radiation.

For this reason, in the subsequent years, all experiments aiming to observe UHECRs were exploiting mainly two approaches: detection of secondary particles at the ground level, and detection of the atmospheric-fluorescence light. In the first category fall big air-shower arrays such as Haverah Park [13], and AGASA [14]. Whilst, in the second one, it is worth to mention the Fly's Eye experiment [15], famous for having observed a cosmic-ray event with an energy of $\sim 3 \cdot 10^{20}$ eV, also known as the Oh-My-God particle [16]. After the successful detection of UHECRs by these experiments, in 1995 an international group of researches started designing the Pierre Auger Observatory (PAO), a *hybrid* cosmic-ray observatory based on both detection techniques mentioned above, where its construction began in 2000.

At the same time, technological advances led in 2003 to the construction of radio-antenna arrays such us LOPES and CODALEMA, that successfully detected radio signals from EASs [17], and [18]. The results obtained

in this field of research, gave birth to the study of radio-emission from EAS at the PAO as well, leading to the installation of the Auger Engineering Radio Array (AERA) in 2010, hence nicely suiting the hybrid conception of the Observatory. Moreover, other enhancements have recently been installed at the Observatory site, i.e. the Auger Muons and Infill for the Ground Array (AMIGA) detector, the High-Elevation Auger Telescopes (HEAT), and the Auger Scintillators for Composition II (ASCII). Thanks to these additional detection techniques, measuring the cascades generated by CRs, it is possible to maximize the amount of information about the primary particles.

In the coming few years, the Pierre Auger Collaboration will push towards a breakthrough in UHECR research; for the first time in history we shall be able to simultaneously measure most of the observables of the primary particle, such as its energy, incoming direction, and muon content, in one shot.

1.3 Thesis Outlook

This thesis presents the investigation of low-energy measurements performed at the Pierre Auger Observatory, by means of the 433 m Surface-Detector (SD) array and AERA. This array has been deployed at the AERA-core site, and it also operates as an external trigger for the radio array. The deployment of this facility has required the optimization of the event-reconstruction process, based only on the collected SD data, as the mutual distance between the surface detectors is a fundamental ingredient for this process. Indeed, the event reconstruction is necessary to infer the energy of the particle that has generated the observed EAS. The CR energy constitutes the building block for the flux spectrum, the starting point of any phenomenological interpretation.

In Chapter 2 we focus on the current interpretation of the UHECR energy spectrum, and briefly on the related astrophysical aspects. In the same Chapter, we give an overview of the fundamental aspects of EASs, and in particular how the CR observables are related to the EAS development in the atmosphere. Chapter 3 presents the Observatory instrumentations, paying particular attention to those providing the data used in this thesis. The exploited methods to extract the high-level information from the raw data are given in Chapter 4, whilst Chapter 5 and Chapter 6 present the analysis based on this reconstructed information. More specifically, Chapter 5 will deal with the energy calibration of the 433 m

SD array, and with the energy-spectrum calculation that follows. Chapter 6 presents an investigation on the sensitivity of the radio measurements that are externally triggered by this SD array and performed by the AERA core. Finally, in Chapter 7 we shall draw the conclusions of this work.

"It is not true that two substances can resemble each other completely and differ only in number"

Gottfried W. Leibniz

"It is presumed that there exists a great unity in nature, in respect of the adequacy of a single cause to account for many different kinds of consequences."

Immanuel Kant

2

Cosmic-Ray Phenomenology

In every-day life one is used to witness recurrent phenomena, such as the sunrise, or apples from trees falling down. One of these facts, is the ceaseless arriving of very high energy subatomic particles entering the atmosphere of the Earth, causing in turn a never-ending *rain* of ionising particles.

The present Chapter deals with the current status of UHECR observations, focusing in particular on the energy spectrum and on our understanding of the air-shower physics that allows us to indirectly study such ultra-high-energy particles.

2.1 Ultra-High-Energy Cosmic Rays

Though the energies of CRs span over more than eleven orders of magnitude, direct measurements can be performed when their energy is below 10^3 GeV. As briefly mentioned in Sec. 1.2, beyond these energies researches may only reckon on indirect measurements, based on the observations of EASs. The reason for such a limitation precisely dwells on the cosmic-ray arrival rate, which depends on their energy. If, for instance, one particle with energy of about 10^2 GeV is expected to arrive *per square meter per second*, we also know that only one particle with energy of $\approx 10^{10}$

GeV will arrive *per square kilometer per year!*

This change in the arrival rate of CRs with primary energy, is described by the differential flux $J(E) = dN/dE$ as the observed number N of CRs with energy between E, and $E + dE$ per unit area, solid angle, and time. The combined cosmic-ray flux from different experiments is shown in Fig. 2.1 versus the energy. The flux spectrum can be approximately described by a power law $dN/dE = E^{-\gamma}$ with a nearly constant spectral index $\gamma \sim 2.7$. However, in what follows we shall focus on the high-energy part of the spectrum ($E \gtrsim 10^6$ GeV), where several features can be noticed.

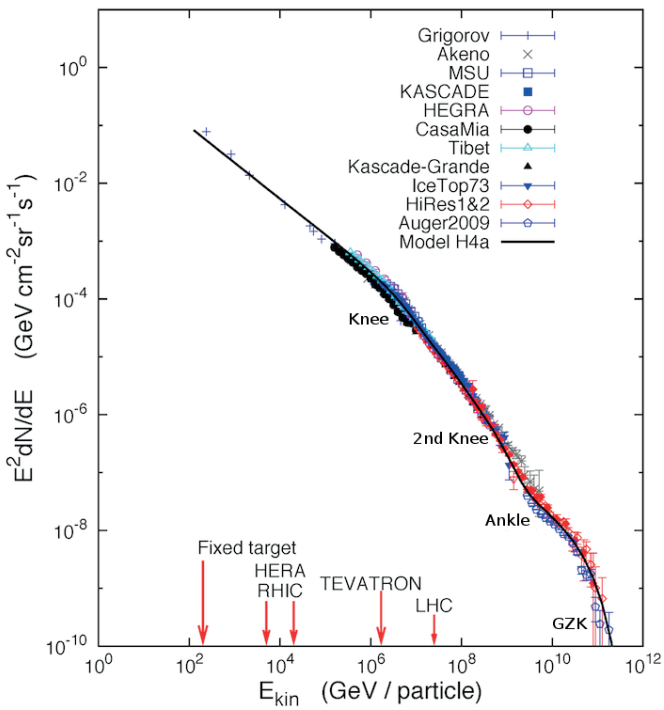

FIGURE 2.1: Scaled flux of cosmic rays $E^2 dN/dE$, as a function of the primary-particle kinetic energy E_{kin}. Image adapted from Ref. [19]

2.1.1 Spectral Features

In Fig. 2.1 the following features are indicated:

i. the *knee:* a steepening of the spectrum at $\approx 3 \cdot 10^{15}$ eV ($3 \cdot 10^6$ GeV), after which the spectral index γ increases to ~ 3.0;

ii. the 2nd *knee:* another steepening of the spectrum at $\approx 5 \cdot 10^{17}$ eV, where γ further increases to ~ 3.2;
iii. the *ankle:* a flattening of the spectrum at $\approx 3 \cdot 10^{18}$ eV, here the spectral index γ decreases to ~ 2.7;
iv. the *cutoff at the end of the spectrum:* an abrupt suppression of the flux of UHECRs at $\approx 6 \cdot 10^{19}$ eV, predicted by Greisen [20], and independently by Zatsepin and Kuz'min [21], in 1966.

Certainly, these features are manifestations of significant changes in one or more characteristics of UHECRs, such a possible change of the composition, propagation-related effects from the source to the Earth, or the suggested different nature of the sources. In the next Sections, some of the most accredited explanations for such changes of the spectral slope will be summarized.

The Dip Region

The energy range including the ankle, approximately from the 2nd knee to the spectrum suppression region, i.e. $\approx 10^{17}$-10^{19}, is called the *dip region*. In recent years, research has been focused on explaining the changes in the slope of the flux spectrum, particularly in this region where a transition from Galactic to Extra-Galactic CRs may take place. Usually, the argument is that no confinement of ultra-high-energy particles is possible by Galactic magnetic fields, and in fact there is no excess in the direction of the Galactic plane. At present, the study of the dip in the UHECR flux is recognized as a key issue in the cosmic-ray-physics research. Actually, it is still unclear where the Galactic Cosmic Ray (GCR) component ends and the Extra-Galactic Cosmic Ray (EGCR) begins, if one accepts the presence of such a transition. In this case, different interpretations of the ankle are compatible with the observations. To properly describe the GCR–EGCR transition, it is necessary to precisely measure the flux and the composition, also in view of the fact that composition measurements are still model dependent.

The presence of the ankle can be interpreted differently. In the following we explain three commonly used models: the *dip model* [22], the *mixed composition model* [23], and the *ankle model* [24]. Fig. 2.2 shows a comparison between the prediction from each of these three models and the measured cosmic-ray spectrum.

The dip model assumes a pure-proton extragalactic component, with

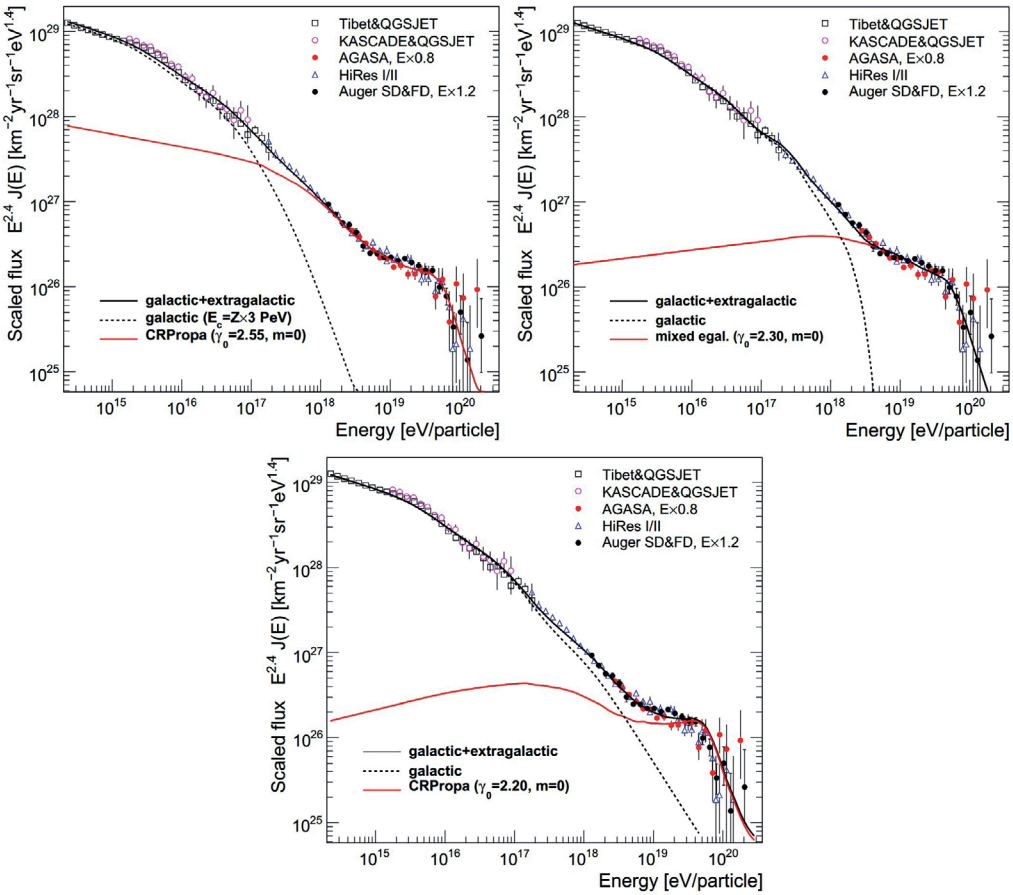

FIGURE 2.2: Predicted flux by models reproducing the GCR–EGCR transition compared to measurements of the cosmic-ray flux spectrum. For the galactic component, $Z \times 3$ PeV is the maximum energy the sources are able to accelerate particles. For the extragalactic component, the CRPropa software [25] parameters γ_0, and m, are respectively the spectral index of the energy spectrum at the source, and the source evolution parameter describing the source density n as a function of redshift z, $dn/dz \propto (1+z)^m$ (so that $m = 0$ corresponds to a uniform source distribution model). *Left panel*: dip model. *Right panel*: mixed composition model. *Bottom panel*: ankle model. Image adapted from Ref. [26].

the GCR–EGCR transition occurring at the 2$^{\text{nd}}$ knee. At the same time, the ankle feature is reproduced as a signature of the proton energy losses through the interaction with the CMB photons, during their propagation from the source to the Earth (cf. Fig. 2.2–*left*).

The mixed-composition model assumes protons with a fraction $\geqslant 15\%$ of heavy nuclei with mass number $A \leqslant 56$. It provides the transition at energies above 10^{18} eV, with the ankle appearing as a signature of the

transition end, whilst the galactic component requires a higher maximum acceleration energy than the previous model (cf. Fig. 2.2–*right*).

Finally, the ankle model provides the presence of the ankle as a natural signature of the transition to the extra-galactic component of a pure proton spectrum, with the galactic cosmic-ray component extending well above 10^{18} eV (cf. Fig. 2.2–*bottom*).

Therefore, different models reproducing this transition have very different implications for the phenomenology of UHECRs and the interpretation of the ankle. Moreover, they give rise to a difference in the distribution of primary CR composition at the dip, and to distinct EGCR injection scenarios at the source, which in turn constrains the species of astronomical objects and/or the acceleration mechanisms. In this respect, a very strong experimental constraint may only be provided from an accurate estimate of the UHECR composition and flux spectrum.

For this reason, the Pierre Auger Observatory deployed several enhancements, such us HEAT and AERAlet, to extend the energy regime of CR observation to lower energies.

The GZK cutoff

The GZK-cutoff prediction concerns the interaction of ultra-high-energy protons, and more generally ions, with CMB photons. The *photohadronic* processes involving a high-energy ion N (with atomic charge Z and atomic mass A) and a 'target' photon, are the following:

- ◇ *photopion production:* $N + \gamma_{CMB} \longrightarrow N' + \pi$,
- ◇ *photopair production:* $N + \gamma_{CMB} \longrightarrow N' + e^+ + e^-$,
- ◇ *photodisintegration:* $N + \gamma_{CMB} \longrightarrow N' + N''$.

The photopion and photopair processes are more important for UHECR protons as energy-loss processes, because the energy threshold for such reactions is higher for large-A nuclei than for protons. On the contrary, photodisintegration is more important for nuclei, where nuclear resonances cause escape of protons, neutrons, and light nuclei from the UHECR nucleus.

Therefore, in case of protons, particles accelerated by distant sources with energies in excess of $E_{\rm GZK} \simeq 6 \cdot 10^{19}$ eV, the so-called GZK limit, can travel a limited distance without interacting. Beyond that range, they

might interact and would continously do so until their energy is below this threshold. Consequently, it is possible to define the GZK *horizon* as the mean distance beyond which UHECRs cannot arrive on Earth without losing energy. The horizon distance decreases with increasing energy, passing approximately from 200 Mpc (Mpc[1]) at $\approx 5 \cdot 10^{19}$ eV to 50 Mpc at energies of about 10^{20} eV for UHECR protons.

These considerations led Greisen, Zatsepin and Kuz'min to argue for the suppression of the UHECR flux at $E \approx 10^{20}$ eV. Although recent results of the Pierre Auger and HiRes observatories does show a suppression in the flux, in agreement with the GZK cutoff [27] [28], there is a number of recorded events with energies $> E_{\mathrm{GZK}}$ that can come from closer sources than 50 Mpc or that could be attributed to heavy composition events.

2.1.2 Origin of Ultra-High Energy Cosmic Rays

The origin of UHECRs continues to be an unsolved problem. Generally, two scenarios have the attention of the scientific community. First, UHECRs are produced by the decay of super-massive *exotic* particles released e.g. from topological defects (possibly created in cosmological phase transitions) or dark-matter candidates. And second, UHECRs are nothing but ordinary particles accelerated by astrophysical sources to extreme energies. As a matter of fact, in the first scenario photons are expected as primaries, which data have indicated to be unlikely [29]. For this reason it will not be considered in what follows, rather a summary on the possible acceleration mechanisms and astrophysical-source candidates is given.

Acceleration Mechanisms

The most plausible theories on cosmic-ray acceleration are based on the Fermi mechanisms, which have firstly been introduced in 1949 by Enrico Fermi [30]. Such mechanisms involve the presence of large-scale magnetic fields in the acceleration region, through which the particles can reach ultra-relativistic energies. Usually, one distinguishes between:

- *first-order Fermi acceleration* where particles are accelerated by strong shock waves propagating through interstellar space,
- *second-order Fermi acceleration* with particles are accelerated by colliding stochastically with magnetic clouds in the interstellar medium.

[1] 1 Mpc $= 10^6$ pc $= 3.26 \cdot 10^6$ ly.

However, as this topic is quite a broad one, going into many details would go beyond the purpose of this thesis. For this reason, it suffices to focus on the first-order Fermi acceleration model. This model is more efficient as it requires only the presence of strong shocked regions, which are quite plausibly present around the expected sources of UHECRs.

According to this model, charged particles gain an amount of energy that is proportional to their current energy per acceleration cycle, where a cycle is the passage of the particle from the non-shocked part to the shocked part of the medium and back. At each cycle, there is a finite probability for the particle to leave the acceleration region, and start its interstellar journey. Therefore, such an acceleration process is gradual, and particles performing as many as possible (irregular) loops in the accelerating region would gain the highest energies. The maximum achievable per-particle energy $E_{\max}$ is governed by the condition that the size L of the accelerating region containing the field must be greater than twice the Larmor radius[2] r_{L} [31]. The Larmor radius of a relativistic particle of charge Ze in a magnetic field B is $r_{\mathrm{L}} = 1.08E/ZB$ pc, where E is the energy of the particle in PeV ($= 10^{15}$ eV) units and B in μG. Taking into account the velocity of the shock front βc, it turns out that L has to be larger than $2r_{\mathrm{L}}/\beta$ [31], and therefore:

$$L \gtrsim \frac{2E}{Z\beta B}\ [\mathrm{pc}] \qquad \Rightarrow \qquad E_{\max} \sim \beta ZBL\ [\mathrm{PeV}]. \tag{2.1}$$

Hence, this equation relates the magnetic field strength and the corresponding size of the acceleration sites, which has been summarized by Hillas [31] for various acceleration-site candidates in his famous diagram shown in Fig. 2.3. It immediately rules out a number of astronomical objects and shows that only few of them (see next Subsection) could act as acceleration sites for UHECRs.

Sources

Keeping an eye upon the Hillas plot in Fig. 2.3, the possible candidates able to accelerate protons up to 10^{20} eV are:

Pulsars Highly magnetized rotating *neutron stars* emitting beams of electromagnetic radiation. These objects are very compact in size and

[2]The radius of the circular motion of a charged particle in the presence of a uniform magnetic field.

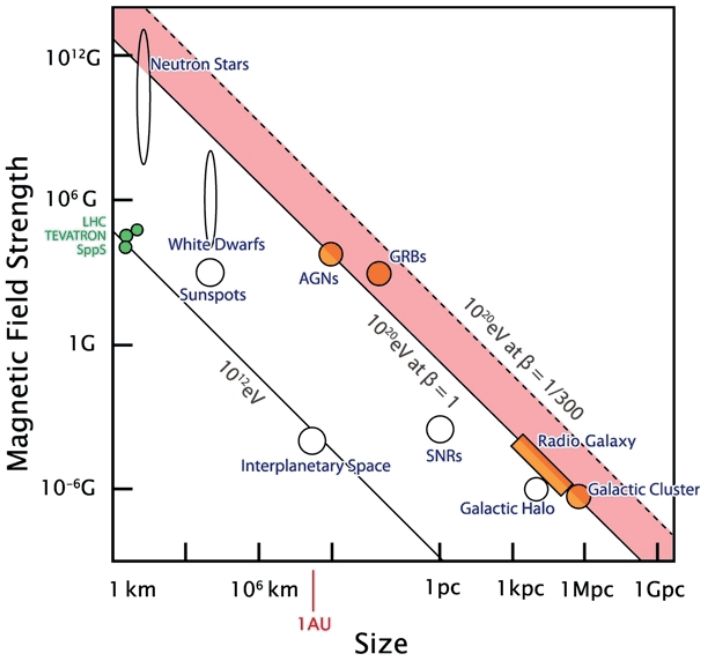

FIGURE 2.3: The Hillas diagram shows the size versus the magnetic field strength of possible astrophysical objects able to accelerate protons up to the indicated energies. The solid and dashed lines indicate respectively $\beta = v/c = 1$ and $\beta = 1/300$ shock-front velocities. Similar lines for iron nuclei would lie below the proton ones. Objects below the diagonal lines cannot accelerate protons to the indicated energies. Image adapted from Ref. [32].

their magnetic field are of the order of 10^{12} G. In this case though, different models could explain the acceleration other than shock acceleration.

Active Galactic Nuclei (AGN) The compact region at the centre of a galaxy whose luminosity exceeds the standard one. Here, magnetic fields are estimated up to 5 G along linear dimensions of 0.02 pc. The emitted radiation from AGN, in so called jets, can be an alternative site of UHECR acceleration, apart from the central region that can contain particles up to 10^{20} eV.

Gamma-Ray Bursts (GRBs) Flashes of gamma rays associated with extremely energetic events, such as collapsing massive stars or merging neutron stars, that have been observed in the Universe originating in jets.

Radio galaxies Types of active galaxy extremely luminous at radio wavelengths. Huge radio lobes are generated by the interaction between

twin jets and the extragalactic medium. The extension of the jets can reach 100 kpc, and magnetic fields at the end of the jets have been estimated to 10 μG, so they can accelerate protons up to $\sim 10^{21}$ eV. The nearby radio galaxy Centaurus A is considered at the moment one of the best candidates for the origin of UHECRs.

Clusters of galaxies Huge objects originating from the clustering of several galaxies by gravitational interactions. Their size can extend up to 500 kpc, and the average magnetic field has been observed to be ~ 5 μG.

Shocks from structure formation Large-scale shocks, with dimension up to 10 Mpc, generated in the accretion flow during structure formation, which can generate magnetic fields of the order of μG.

Propagation

The task of understanding the origin of UHECRs is further made difficult by their propagation through the interstellar medium. Here, the presence of magnetic fields and radiation fields introduces additional processes that, depending on the cosmic-ray energy, affect to a certain extent the measured flux spectrum on Earth. If on one side the interaction with the CMB and the CIB might cause the flux cutoff (see Sec. 2.1.1 above), on the other side, the interaction with the magnetic fields causes a deviation of cosmic-ray trajectories so that the arrival direction does not point back to their source. As a result, the arrival-direction distribution (cf. Fig. 1.2) of UHECRs shows no clear evidence of anisotropy [3].

The study of UHECR propagation across the interstellar space, taking into account all the mentioned effects above, continuously challenges our understanding of the processes involved, and in general of the Universe as a whole.

2.2 Extensive Air Showers

An EAS is an avalanche of particles generated by the entrance of a subatomic particle in the Earth's atmosphere. This avalanche starts as the primary CR, i.e. the particle which has travelled from "somewhere" to the Earth, undergoes a nuclear interaction with an air molecule, and several *secondary* particles are created as products of this *first interaction*. In turn, each secondary continues to interact in the atmosphere, thus gener-

ating other particles that eventually continue to generate more and more particles either by interaction or by decay processes. At the same time, because the energy of the primary CR gets distributed among a swarm of secondaries, this cascade-generating process does not continue indefinitely, and once the available energy per produced particle drops below a certain critical value, ionisation energy-loss processes take over and the number of particles in the shower start to decrease.

In other words, the entering of CRs into the atmosphere gives rise to a number of processes that all together depict an air shower: an ensemble of *causally-connected* particles, moving at approximately the speed of light, increasing and decreasing dynamically in number as the shower develops. With increasing energy of the primary cosmic ray, the number of secondary particles taking part in the shower will increase as well; but, the *type* of secondary particles that can be created depends on the *nature* of the primary CR, i.e. on its chemical composition.

Most primary CRs are familiar stable subatomic particles that normally occur on Earth, such as protons, atomic nuclei, photons (gamma rays), electrons[3] and neutrinos. The main difference between air showers initiated by different primaries, is the difference in the density of the various secondary particles, i.e. how many particles of a given type are present in the shower. Loosely speaking, it is possible to identify three major *components* interplaying in an EAS (cf. Fig. 2.4):

electromagnetic (soft) component: electrons, positrons and photons, created via repetitive electron-pair creation and bremsstrahlung processes.

hadronic (hard) components: protons, neutrons, mesons (such as pions and kaons), and diverse nuclear fragments created by the leading-baryon emerging from the first interaction.

muonic component: muons and muon neutrinos created from meson decays.

As mentioned above, the nature of the primary CR plays an important role in the population of these components. A gamma-ray-initiated air shower will create an *electromagnetic cascade*, as electron-pair production takes place in the field of atmospheric molecules due to momentum conservation. The resulting electron and positron would then produce sec-

[3] In general, no significant contribution to the rate of detectable EASs is expected from cosmic-ray electrons, due to the little chance they have of travelling far distances because of synchrotron and inverse-Compton energy losses.

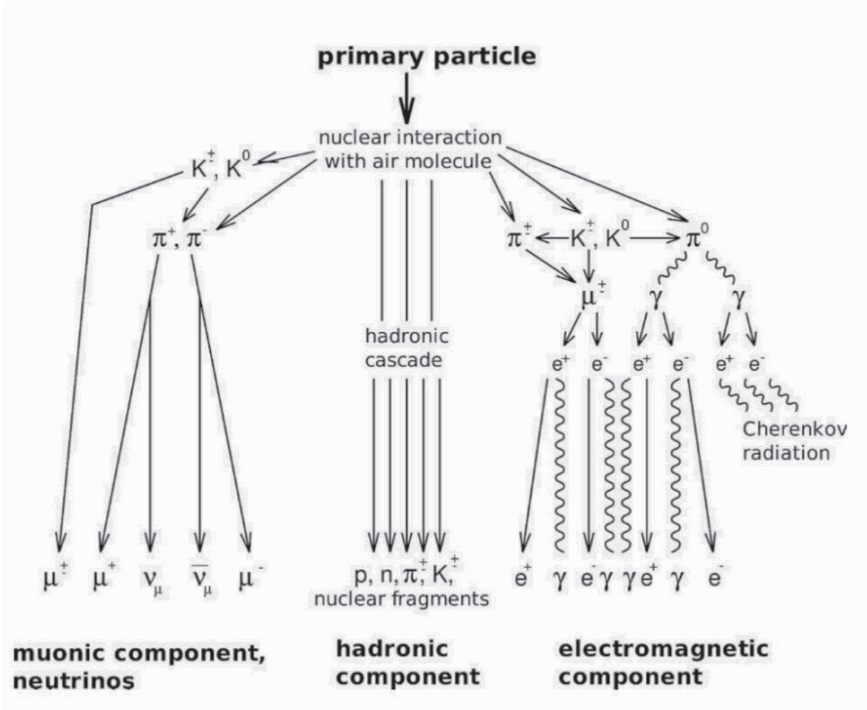

FIGURE 2.4: Schematic representation of a developing cascade. Image taken from Ref. [33]

ondary gamma rays by bremsstrahlung, which in turn would produce new pairs, further producing bremsstrahlung photons. This exponential growth of the cascade, populates mainly the soft component, and would continue before the critical energy $\epsilon_\gamma \approx 1$ MeV is reached.

Conversely, as a nucleon interacts with an air nucleus a 'leading baryon' emerges from the first interaction along with several mesons, mainly kaons, charged and neutral pions. On the one hand, the leading baryon feeds the hard component of the shower, a hadron cascade that propagates along the direction of the initial momentum vector of the incident primary CR. On the other hand, the charged mesons continue to interact, producing other charged and neutral mesons, until their energy is reduced to ≈ 20 GeV, when decay to muons and muon neutrinos start to dominate and the muonic component starts to be populated. However, for neutral pions, the decay to two photons is dominant at all energies, because of their short lifetime; consequently each produced high-energy photon would initiate an electromagnetic cascade, populating in this way the soft component.

Although the difference between a gamma-ray and a single-nucleon (a

proton for instance) initiated showers seems clear at least in terms of shower-component populations, the same cannot be said when a hadronic shower would be initiated by a nucleon or by a heavy nucleus. As the primary particle is a heavier nucleus of mass number A and energy E the usual picture of the resulting shower is a superposition of A proton-induced showers of primary energy E/A. Therefore, it would be very hard to distinguish, only in terms of shower components, an air-shower generated by an A-nucleus of energy E from an air shower initiated by a single-nucleon of energy E/A.

Effectively, as we are going to see in the next Section, there exist other key factors that characterize a given air shower: the longitudinal evolution of the *number* of produced secondary particles in the atmosphere, and the lateral change of this number with respect to the central part of the shower.

2.2.1 Dynamics of Air-Shower Development

First, let us define a few primary-type-independent concepts that simplify the following discussions, illustrated schematically in Fig. 2.5:

zenith angle θ: is the angle of incidence of the primary CR with respect to the vertical ($\theta = 0$);

shower axis: the direction of motion of the primary CR initiating the air shower;

shower core: the central region of an air shower around the shower axis, where the particle density is highest;

shower front: the particle disk of finite thickness where all the produced secondaries lie;

atmospheric depth X [g $\cdot$ cm^{-2}]: is a measure of the amount of traversed atmospheric matter per unit area in the vertical column of air, by an air shower from its top ($X = 0$). For inclined trajectories, one uses the expression ***slant depth*** $X_s \approx X \sec\theta$ [g $\cdot$ cm^{-2}].

These defined quantities will be used all along this thesis, as they allow to further define all other variables and parameters related to the atmospheric cascade development, and ultimately to the physics governing the dynamical evolution of EAS.

Longitudinal Profile

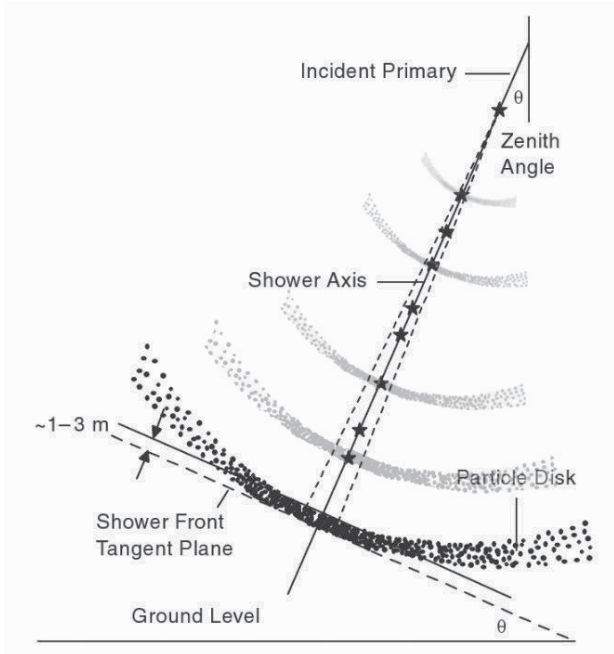

FIGURE 2.5: Schematic representation of a developing air shower in the atmosphere, and the corresponding shower-related quantities. Image adapted from Ref. [34]

The development of an EAS in the atmosphere is actually not simple and easily explained, as it emerges from the secondary-particle dynamics. As already mentioned, various types of particles can be created and each type of particle interacts differently in the atmosphere. In addition, given a particle of a certain kind, its kinetic energy determines whether this particle will continue to generate new particles, or rather will be absorbed.

Therefore, it is certainly easier to identify many-body properties of an EAS, the most straightforward being nothing but the *number* $N(X)$ of produced particles at a certain atmospheric depth X, called *longitudinal shower profile*. The longitudinal profile is mainly a function of the primary CR energy E_0, the angle of incidence θ, and the *height of the first interaction* h_1 above sea level. Instead of the height h in the atmosphere, it is usually more convenient to use the atmospheric depth X, related to the first through a formula that in turns depends on the atmospheric density. As this density is not constant with altitude and location on Earth, one needs an atmospheric model (whose parameters depend on the specific location) to perform such a conversion. Hence, in what follows, the atmospheric depth will be used to specify locations along an air-shower

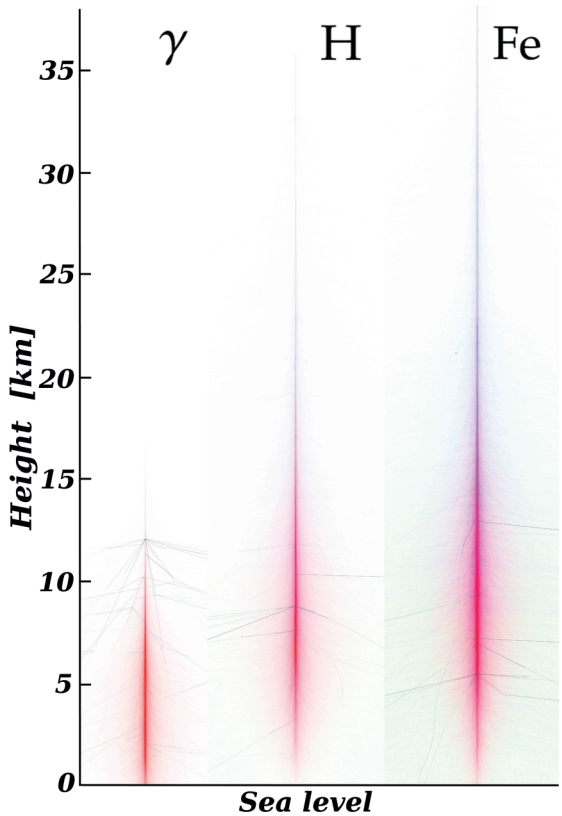

FIGURE 2.6: Schematic representation of three vertical ($\theta = 0^\circ$) EAS respectively generated by a photon (left), a proton (center), and a iron nucleus (right), at an energy of 10^{17} eV by means of the CORSIKA simulation tool [35].

track in place of the altitude, as this quantity includes by definition this height–atmospheric density dependency.

Though the longitudinal profile does not depend strongly on the primary mass A, the location of its maximum $N_{\max}$ does depend on the primary CR composition. This location in the atmosphere is widely known as *depth of maximum development* $X_{\max}$, where one has $N(X = X_{\max}) = N_{\max}$. Indeed, given a primary energy the depth of the first interaction X_1 decreases with increasing primary mass A, because the chance of interacting (interaction cross-section) increases with the projectile mass. As an example, three vertical ($\theta = 0^\circ$) EASs have been simulated for three different types of primary particles, respectively a γ, a proton, and an iron nucleus all having a primary energy of 10^{17} eV. The simulations have been performed by means of COsmic Ray SImulations for KAscade (CORSIKA) [35], using SIBYLL as the high-energy hadronic inter-

action code [36]. Fig. 2.6 shows the trajectories across the atmosphere of the secondary particles of energy $\geqslant$ 3 GeV, down to the sea level, in a two-dimensional representation obtained with COrsika dAta accesS Tools (COAST) [37]. As it can be seen, the γ primary starts to interact very deep in the atmosphere compared to the hadron primaries, and the proton deeper than the iron nucleus, though not that much. Fig. 2.7 shows the longitudinal profiles for these air showers, where the same behaviour can be observed for $X_{\max}$ as well.

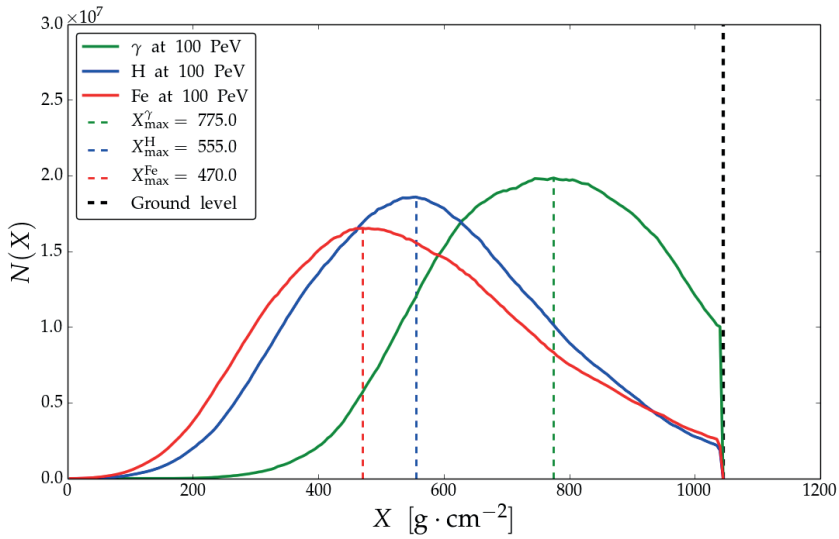

FIGURE 2.7: The longitudinal profiles of the three vertical ($\theta = 0^\circ$) air showers respectively generated by a photon (solid-green line), a proton (solid-blue), and a iron nucleus (solid red), with the same primary energy of 10^{17} eV. The $X_{\max}$ position is also indicated by the dashed lines.

However, since the primary-mass dependency of $X_{\max}$ is ultimately related to the interaction cross-section, i.e. to a probability for the CR of interacting at "some point", even with identical particles having the same energy and zenith angle will have different longitudinal profiles, and consequently, different $X_{\max}$. Fig. 2.8 shows, as an example, these *shower-to-shower fluctuations* for the three simulated air showers showed above. These fluctuations are also due to the several stochastic processes that compete in the development of air showers, as a consequence of the low density nature of the atmosphere. To mimic these effects, for each different primary, a CORSIKA simulation has been launched ten times, and every time a different random seed has been chosen for the high- and low-energy interaction models. As the interaction cross-section is directly

proportional to the projectile atomic mass, one observes smaller fluctuations for heavy nuclei than for light ones. In addition, hadronic processes are the main contributors to air-shower fluctuations, and purely electromagnetic showers are in general less subject to fluctuations.

The longitudinal profile of charged particles composing an air shower can be directly measured by optical fluorescence detectors, which are sensitive to the fluorescence light emitted by nitrogen molecules in the atmosphere after the passage of an EAS (see Sec. 3.2). At the same time, electromagnetic pulses are caused by the changing number of secondaries charge carriers, which makes the radio-detection technique sensitive to the derivative of $N(X)$. On the contrary, surface arrays measure the particle density at ground, i.e. at the last stage of the longitudinal profile $N(X_{\text{ground}})$. As we shall see in the next Section, the air shower at the ground level is characterized by the *lateral displacement* of the particles in the shower front with respect to the shower core.

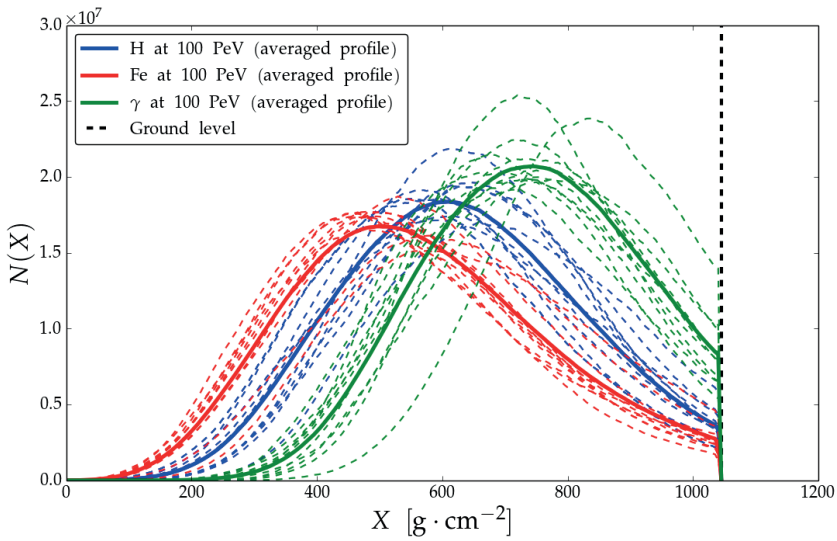

FIGURE 2.8: Illustrative example of shower-to-shower fluctuations in the longitudinal profiles of the three vertical ($\theta = 0°$) air showers in Fig. 2.6 respectively generated by a photon (green lines), a proton (blue), and a iron nucleus (red), with the same primary energy of 10^{17} eV. Each dashed line corresponds to a simulated air-shower longitudinal profile, where only the random seed has been changed, for every primary. The solid lines represent the average longitudinal profile for each primary.

Lateral Distribution

Together with the longitudinal profile, the lateral structure of the shower front, i.e. the *Lateral (density) Distribution Function (LDF)* of the particles, characterizes air showers and is used to compute the particle content N, required in turn to estimate the primary energy.

As for the longitudinal profile, the lateral spread emerges as a many-body feature of the particle dynamics responsible of the EAS development. Indeed, already at the first few hadronic interactions, the secondary particles produced inherit a *transverse momentum* component by momentum conservation. In the same way, the resulting momentum of muons produced by hadron decays partly gains this transverse component, further enhancing the lateral spread. At the same time, photon-initiated electromagnetic sub-cascades develop through different interactions, i.e. pair production, bremsstrahlung, Compton and multiple Coulomb scattering, excitation and ionization, which in turn increase the lateral displacement of electromagnetic particles.

Therefore, the shower-front lateral structure of an EAS is again resulting as a superposition of competing mechanisms of hadronic and electromagnetic nature. For this reason, the LDF of a purely electromagnetic cascade is essentially different from a hadronic one, as the muonic-component contribution is almost absent from the first. As an example, Fig. 2.9 shows the two-dimensional LDF for the three vertical air showers corresponding to the simulated events in Fig. 2.6 and to the longitudinal profiles in Fig. 2.7. As it can be seen, the photon-initiated shower presents an LDF (top-left panel) spatially very compact i. e. the bulk of secondary particles is distributed close to the shower core and it dramatically drops after few hundreds meters (the purple area indicates an extremely low particle density). The lateral extension of the profile for an electromagnetic shower is directly related to the atmosphere Molière radius [38]. On the contrary, for the proton-initiated shower the LDF (top-right panel) extends to larger distances, where the muonic components starts to have the same weight as the electromagnetic component. Finally, for the iron-initiated shower the LDF (bottom panel) appears to be more extended than the proton shower due to the larger number of muons, though the general behaviour is basically the same with a smooth drop as the distance increases.

The all-particle density $\rho(r)$ is in first approximation axially symmetric, and for this reason usually is only a function of the radial (lateral) distance r from the shower axis. Asymmetries in the particle density

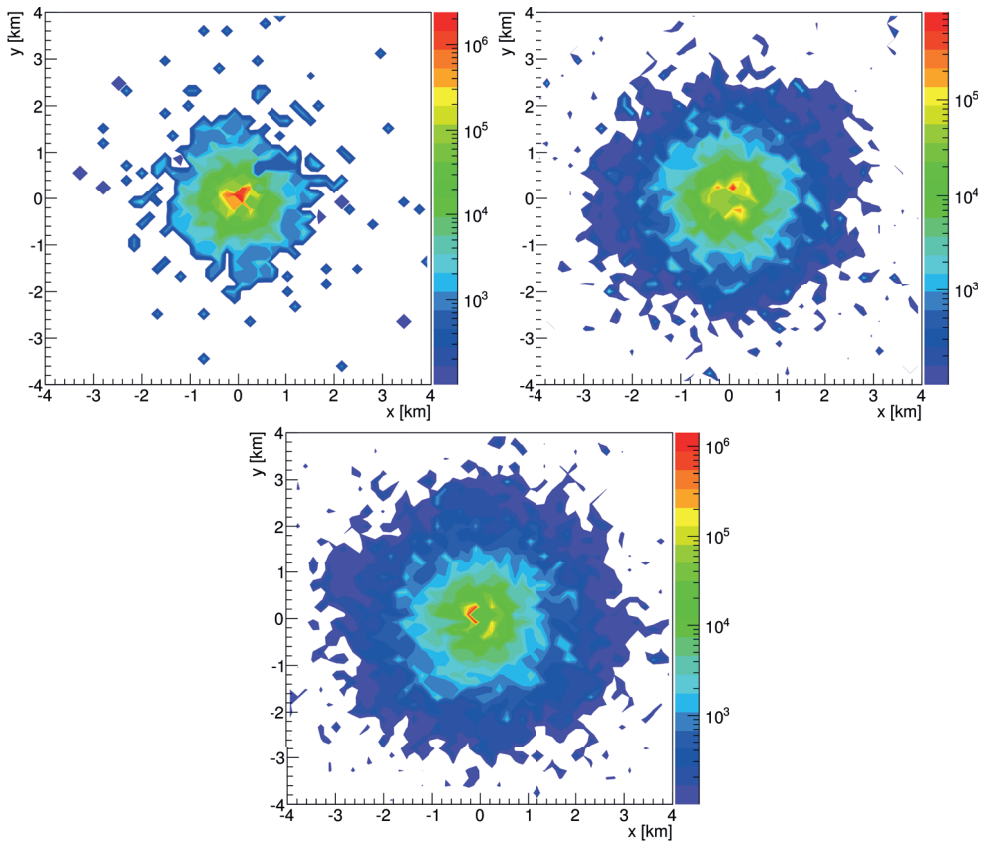

FIGURE 2.9: Illustrative example of two-dimensional LDF (at 1400 m above sea level (a.s.l.)) for the three vertical ($\theta = 0^\circ$) air showers in Fig. 2.6 respectively generated by a photon (top left), a proton (top right), and a iron nucleus (bottom), with the same primary energy of 10^{17} eV. The color scale indicates the all-particle density in arbitrary units.

arise from zenith-angle and geomagnetic effects, which become more important for strongly inclined showers, and act differently on the muonic and electromagnetic components. Though along this thesis only *vertical* ($\theta \leqslant 55^\circ$) air showers will be selected, there is a zenith-angle effect that cannot be neglected.

This effect is related to the increasing amount of atmosphere traversed by air showers with increasing zenith angle, and it is known as *atmospheric attenuation* (see e.g. [39]): as the shower crosses more atmosphere more particles will be absorbed. In addition, the rate of absorption ultimately depends on the particle type, with the final effect that the particle mix changes as well with increasing zenith angle. As a consequence, at sea level, the electromagnetic component of inclined showers will be

greatly absorbed, while the muonic component is hardly affected.

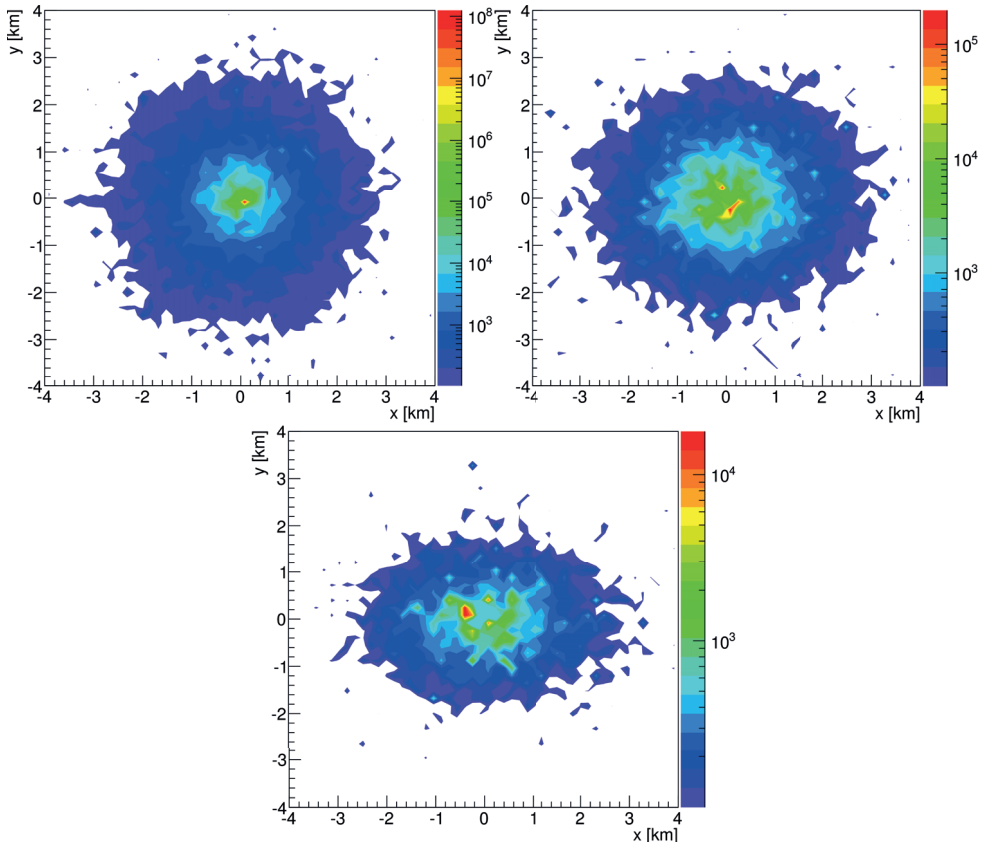

FIGURE 2.10: Illustrative example of two-dimensional LDF for three iron-initiated air showers respectively simulated with $\theta = 0^\circ$ (top left), $\theta = 30^\circ$ (top right), and $\theta = 50^\circ$ (bottom), with the same primary energy of 10^{17} eV, and azimuth angle of 180°. The color scale indicates the all-particle density in arbitrary units.

Therefore, depending on the observation level X, the atmospheric attenuation of the shower particles affects, to a certain extent, the observed LDF. In addition, as the total number of particle $N(X)$ (which is proportional to the primary CR energy) is estimated by the (integrated) particle-density distribution, two showers differing only for their zenith angles θ will *appear* with different energies, if this effect is not taken into account. In this case, the difference in the particle content, which is then reflected in the LDF normalization factor, has to be attributed to the particle absorption rather than to a different primary energy. To illustrate this point, Fig. 2.10 displays an iron-initiated shower simulated three times keeping the energy constant at 10^{17} eV, and each time only changing the zenith angle ($\theta = 0^\circ, 30^\circ, 60^\circ$). As it can be noted, though the primary

energy is fixed, the particle-density scales inversely with the zenith angle (cf. the colour scale in this Figure). This effect is more important than the azimuthal-asymmetry feature, also visible in this Figure, arising mainly for geometrical reasons, which mostly influences the tail of the LDF.

To correct for this effect, the Constant Intensity Cut (CIC) method is commonly used in the analysis of ground-array air-shower data (see e.g. [40]). Basically, it relies on the assumption that the primary CR flux is isotropic, and consequently the arrival rate of air showers generated by primaries with the same energy and composition is constant as a function of the zenith angle at a given energy [41]. Therefore, the observed change in the rate is related to its zenith-angle dependency. In Section 5.3.1, we shall apply this method to find the corrected energy estimator from which one can determine the primary CR energy.

Radio Emission

The electrons, positrons and photons forming the soft component of an EAS are more than 90 % of the total number of secondary particles created in the cascade. Indeed at each hadronic interaction a sort of "energy feeding" takes place from the hadronic component to the electromagnetic component of the air shower, until the energy of the charged pions is low enough and decays are favored over interactions. In the present Section, we shall see how their intricate dynamics leads to pulsed emissions of electromagnetic waves in the radio frequency band.

During the development of an EAS, electrons and positrons are clustered in a plasma cloud that moves downwards with almost the light velocity. In the course of its motion through the atmosphere, the charges will experience the geomagnetic field $\mathbf{B}_{\oplus}$ by means of the Lorentz force, which will deflect oppositely charged particles to opposite directions. The Lorentz force will induce the charges in the cloud to move with a drift velocity perpendicular to the shower axis. This drift velocity, results to be opposite for electrons and positrons, causing the rise of a continuous flow of a time-varying transverse electric current in the shower front, proportional to the varying total number of charges $N_{\mathrm{e}}(X)$. Each time a new generation of electron-positron pairs is created, charge separation will occur all over again. Therefore on average, after many pair-production processes, one has a quasi-steady state in which negative and positive charges are localized in opposite regions of the cloud, roughly at a constant lateral separation, moving in parallel to each other: as a matter of fact a 'moving electric dipole' comes in addition at the shower

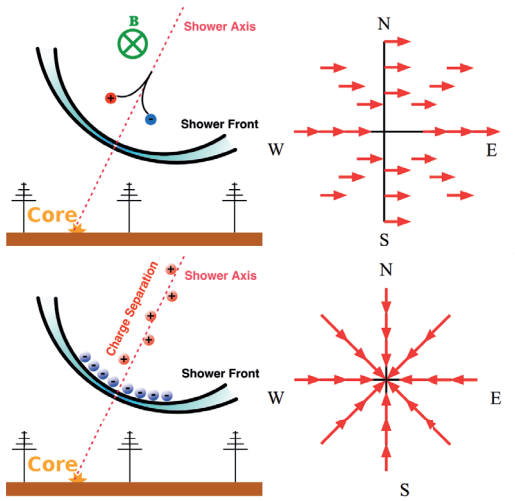

FIGURE 2.11: Schematic illustration for the two main radio-emission mechanisms occurring during the development of an EAS. *Top panel:* Charge separation induced by the geomagnetic field, together with e^- and e^+ creation/absorption, generate a time-varying transverse current, which constitutes the main contribution. *Bottom panel:* the Askaryan effect, responsible for a time-varying negative charge-excess, represents the second important contribution for the total radio emission (see text for details). Image adapted from Ref. [42]

front [43, 44, 45, 46, 47, 48].

At the same time one has to keep in mind that the charges are moving in a dielectric medium, that in turn consists of a huge number of molecules. Here, at the passage of the air shower, less energetic secondary particles are being slowed down. The action of the geomagnetic field is to pull apart positive and negative charged particles and to leave behind a 'static electric dipole' distributed along the path of the shower at its passage.

Moreover, depending on the energies involved, other processes can take place: as the energy of secondary particles drops at $\approx$ 30 MeV, electrons in the cloud may knock-out orbiting electrons from the molecules of the atmosphere, positrons annihilate in flight and photons drag orbiting electrons by Compton recoil. These three processes are responsible for the arising of a net excess in the number of electrons in the air shower [49], known as charge excess, which depending on the per-particle energy considered, varies from $\sim$ 20 % for secondary particles of $\approx$ 2 MeV to $\sim$ 1% for particles of energy $\approx$ 100 MeV [50].

Fig. 2.11 shows a schematic illustration of the two main competing emission mechanisms from EASs, though it is widely accepted that the main emission mechanism contributing to the radio emission is the geomag-

netic one [51, 52, 53]. Such an emitted radiation is linearly polarized and the electric field vector $\mathbf{E}_{\mathrm{G}}$ is oriented in the direction of the Lorentz force, perpendicular to the shower axis, and independent from the relative orientation of the observer with respect to the air shower. This geomagnetic emission is coherent when the wavelength of the out-coming radiation is comparable with the thickness of the charged cloud, and it scales with the sine of the opening angle α between the shower velocity $\mathbf{v}$ along the axis and $\mathbf{B}_{\oplus}$

$$\mathbf{E}_{\mathrm{G}}(\boldsymbol{x}, t) \propto -\mathbf{v} \times \mathbf{B}_{\oplus} \propto \sin\alpha \tag{2.2}$$

The second important contribution is the aforementioned charge-excess one, also known as the Askaryan effect, leading as the shower develops to a time-varying net charge. The relative strength of this contribution has recently been measured to be on average $14 \pm 2\%$ (at the AERA site) [54], by means of polarization measurements. Indeed, the charge-excess emission has a different polarization signature, the electric field vector $\mathbf{E}_{\mathrm{C}}$ is oriented radially with respect to the shower axis, so its orientation will depend on the observer position. As a consequence, the polarization of the total electric field $\mathbf{E}(\boldsymbol{x}, t)$ will depend as well on the position of the observer.

In addition, the air shower and the emitted radiation propagate both in the same medium: the atmosphere, which has an altitude-dependent refractive index slightly larger than unity. As a consequence, radio pulses propagate slower than the air shower in the atmosphere, ultimately experiencing Cherenkov-like time compression with the effect of enhancing the radio emission at certain observer positions [47, 55]. Such an effect is particularly relevant at higher frequencies ($\gtrsim$ 300 MHz), and although this affects both emission mechanisms, it does not affect the polarization direction of the emitted radiation.

For the polarization analysis in Chapter 6, it is then possible to consider the total electric field as the vectorial sum of the geomagnetic and of the charge-excess radio-emission contributions:

$$\mathbf{E}(\boldsymbol{x}, t) = \mathbf{E}_{\mathrm{G}}(\boldsymbol{x}, t) + \mathbf{E}_{\mathrm{C}}(\boldsymbol{x}, t). \tag{2.3}$$

To describe the polarization state of electromagnetic radiation, we use the

Stokes formalism defining the parameters (see e.g. [56]):

$$\begin{aligned} I &= \langle |E_x| \rangle^2 + \langle |E_y| \rangle^2 && \text{the total intensity of the signal,} \\ Q &= \langle |E_x| \rangle^2 - \langle |E_y| \rangle^2 && \text{linear polarizations,} \\ U &= \langle 2\mathrm{Re}(E_x E_y^*) \rangle && \text{diagonal linear polarization,} \\ V &= \langle 2\mathrm{Im}(E_x E_y^*) \rangle && \text{right-handed and left-handed circular polarization.} \end{aligned} \tag{2.4}$$

Here, $\mathbf{E} = \mathbf{E}(t)$ is the total electromagnetic field, x and y are defined as the coordinate axis in the plane perpendicular to the propagation vector of the electromagnetic wave, and $\langle \ \rangle$ indicates the time average. It is easy to see that this formalism is extremely convenient as any polarization state can be described as combination of linearly and circularly polarized components, with degree of polarization 0 (unpolarized) $\leqslant p \leqslant$ 1 (completely polarized) given by:

$$p = \frac{\sqrt{Q^2 + U^2 + V^2}}{I} \tag{2.5}$$

The renewed interest in the radio detection of cosmic rays, has led to the development of detailed models describing the radio-emission mechanism. It turns out that two different ways of approaching the problem have given rise to conceptually different models. Indeed a formal description of the phenomenon can not be separated from our point of view in looking at the physical process; namely macroscopic phenomena can be understood by looking at the individual behaviour of its microscopic constituents and vice versa microscopic many-body systems can be described through macroscopic variables that reflect well the collective behaviour of the whole system.

Here we have a quite common situation in which the collective dynamics of a tremendous amount of 'microscopic' particles arises in a 'macroscopic' emission of electromagnetic waves. If in a microscopic approach the full emission is obtained adding up the emissions for each particle trajectory in the shower (microscopic models can be found e.g. in [57, 58, 59]), in a macroscopic approach the relevant particle distributions and currents in the shower are used as input in the Maxwell equations to calculate the electromagnetic radiation (see for example [43, 47]).

It has been demonstrated that the two approaches provide equivalent descriptions of the same physical process (see e.g. [60]), the CORSIKA based CoREAS [61] is widely used in the community to simulate the radio emission from EASs. As an example, Fig. 2.12 shows the simulated

power of the radio signal with CoREAS (filtered between 30 -80 MHz) eventually observed at 1400 m a.s.l. in the shower-plane reference frame, where the vertical axis is aligned along the propagation vector of the radio wave (cf. Fig. 4.4). The radio-signal footprints have been obtained from a proton- (*left panel*) and an iron-initiated (*right panel*) vertical ($\theta = 0^\circ$) air shower at 10^{17} eV. For comparison, the particle density at the surface of the Earth is shown in Fig. 2.9-*top-right* for the proton, and Fig. 2.9-*bottom* for the iron.

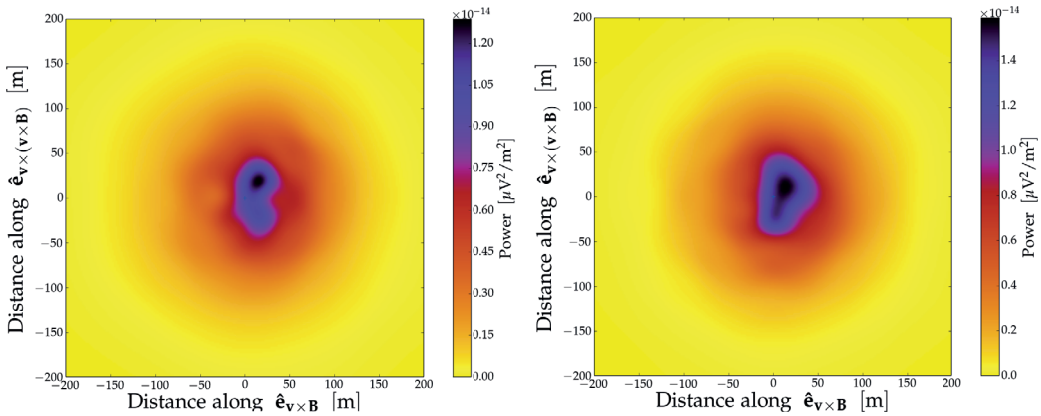

FIGURE 2.12: Simulated footprints of the radio signal for a proton- (*left panel*), and an iron-initiated (*right panel*) vertical ($\theta = 0^\circ$) air showers, with the same primary energy of 10^{17} eV. In calculating the power, the signal is filtered between 30 - 80 MHz, which is the frequency bandwidth used by the AERA antennas at Auger (Section 3.4). The observation level is 1400 m a.s.l., the height of the Pierre Auger Observatory.

2.3 Final Remarks

Our understanding of UHECRs is essentially connected to our understanding of EASs, as only indirect measurements of EASs reveal us the precious information carried by these cosmic messengers. Recently, our understanding of physical phenomena is validated by their reproducibility by means of simulations, which are a fundamental tool in our hands to further improve this understanding. As mentioned in Section 2.2.1, some EASs originate from hadronic interactions at energies beyond those reachable in man-made colliders. As a matter of fact the highest-energy measurement[4] of the proton-air cross-section has been performed by the

[4]center-of-mass energy per nucleon of 57 TeV equivalent to a CR energy of $\approx 2 \cdot 10^{18}$ eV.

Auger collaboration [62]. In addition, as EASs develop in the atmosphere, the condition and the characteristics of the atmosphere have been deeply investigated (see e.g. [63]), and atmospheric models need to be included in any reliable simulation of EASs.

Then, it is easy to see that hadronic-interaction models are a key element of the EAS simulations. Unfortunately, especially at highest energies ($\gtrsim 10^{19}$ eV) and for inclined events, the predicted number of muons is smaller than the observed one [64, 65]. In addition, although the interaction models have been recently tuned to LHC results, none of them is able to deliver a consistent prediction of both the electromagnetic and muonic components as observed by Auger. This muon deficit is a further reason to understand high-energy physics, and UHECRs furnish once again the ideal "laboratory" to test our current knowledge.

"It is beyond a doubt that all our knowledge begins with experience."

Immanuel Kant

3

The Pierre Auger Observatory

The present Chapter gives an overview of the different detection techniques deployed at the PAO, from its birth to the present time, including a brief summary of the foreseen extensions in the near future.

The PAO is the largest UHECR detector in the world, and it is located on the vast plain of the Pampa Amarilla at an altitude of 1400 m, close to Malargüe in the Mendoza Province, Argentina. The choice of this location has been the result of a site survey all around the globe, including Australia, South Africa, Spain, Russia, and the United States. In the decision, physics considerations have mainly motivated the selection criteria, in particular, the flatness and accessibility of the surface-detector-dedicated area, the altitude, and the good visibility for the fluorescence-light observation at the site [66].

The PAO is the first experiment that has combined both ground detectors and fluorescence detectors at the same site. More than that, the Observatory hosts additional air-shower-detection techniques. On the one hand, radio pulses from cosmic-ray-induced air showers are detected with the aid of AERA. On the other hand, specific muon detectors have been installed to discriminate the muonic component from the electromagnetic component.

3.1 A Hybrid Concept

As mentioned above, the uniqueness of the PAO derives from its hybrid design (which nowadays includes four detection techniques): the integrated surface array and the fluorescence detector system. The ground-detection area amounts to about 3000 km^2, where 1600 Surface Detector (SD) stations are placed on a triangular grid with mutual distance of 1.5 km between them. At the edge of the surface array, there are 24 Fluorescence Detector (FD) telescopes, grouped by six in four buildings (Coihueco, Los Leones, Los Morados, and Loma Amarilla), overlooking the atmosphere above the observational area (cf. Fig. 3.1).

This original design was proposed to allow the detection of EAS generated by UHECRs of energies beyond 10^{18} eV. However nowadays, the low-energy upgrades AMIGA and HEAT are operational as well, and the lower limit of $\sim 10^{17}$ eV has been reached recently. The AMIGA extension consists of an array of SD stations, covering about 24 km^2, placed at a mutual distance of 750 m, together with additional underground muon detectors in the same area. The HEAT extension consists of three fluorescence telescopes, close to Coihueco, looking at a higher part of the atmosphere than the other telescopes, where the shower profile of low-energy events can be recorded.

The observation of EASs crucially relies on the communications between the Central Data Acquisition System (CDAS) and the different detectors of the Observatory disseminated over a large area. The CDAS performs all the operations concerning the identification of the event triggers and the collection of the data, by sending commands to and receiving local information from the telescopes and the individual detectors in the field.

The SD array is divided into four sectors associated with the four FD sites, where the communication towers are installed. The individual tanks communicate through a custom-made Wireless Local Area Network (WLAN) to these communication towers. At the same time, a high capacity microwave network guarantees the communications between the four FD sites, the CDAS and the control center located at the campus building in Malargüe [67].

3.2 Fluorescence Detector

As mentioned in Section 1.2, one of the techniques used to detect UHECRs is the collection of *air fluorescence emission* due to the passage of

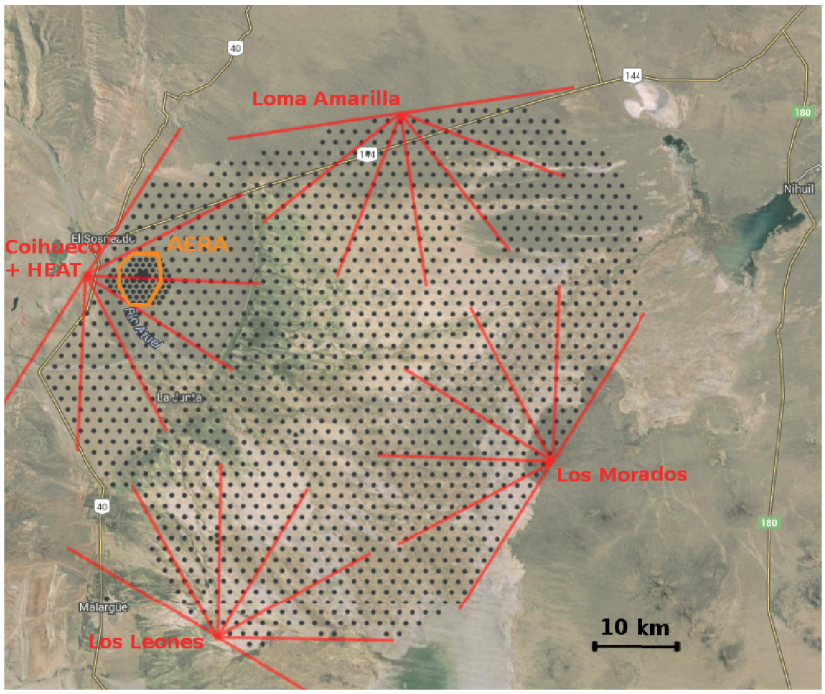

FIGURE 3.1: Schematic layout of the Pierre Auger Observatory overimposed on a satellite image of the actual site close to Malargüe, in the province of Mendoza, Argentina. Each black dot indicates an SD tank, whilst the red lines give an idea of the field of view associated with the FD telescopes at the perimeter of the surface array. The Infill (the more dense area of black dots) and AERAlet (the most dense in the middle of the Infill) arrays are located close to Coihueco (to the west), where the AERA antennas are placed as well.

particles in the atmosphere. The charged particles produced along the shower track, lose their energy by inelastic collisions with the nitrogen and oxigen molecules in the atmosphere, which are excited. The nitrogen molecules isotropically emit fluorescence light, in the $\sim$ 300 - 430 nm range, to return to their ground state. The number of emitted photons is proportional to the energy deposited by the charged particles through electromagnetic energy losses [68]. Therefore, measuring this light at many atmospheric depths, an FD is able to map the *longitudinal development profile* $dE(X)/dX$. The integral of the longitudinal profile is nothing but a calorimetric measure of the total electromagnetic shower energy, which is $\approx$ 90% of the primary energy of the CR. The unmeasured energy is mainly dissipated in creating muons and neutrinos, and its amount depends on the primary CR composition. On average, this *missing* energy is about 5% for proton-induced showers and 15% for iron-induced show-

ers [69].

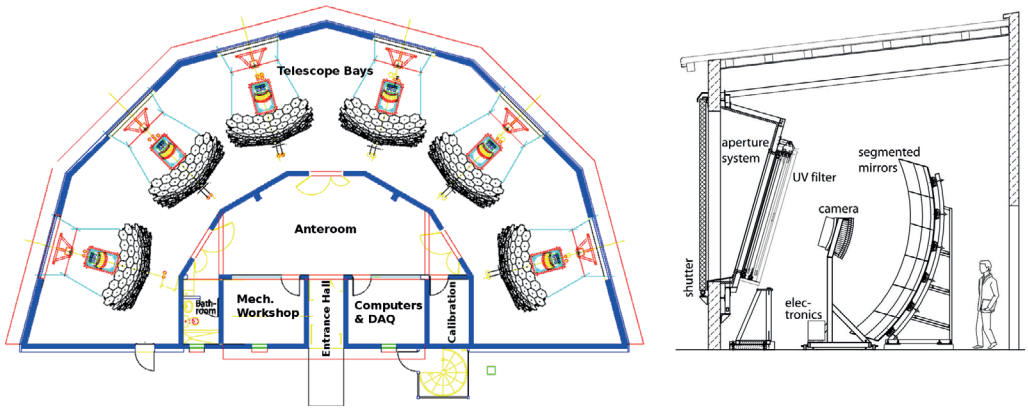

FIGURE 3.2: *Left panel:* illustration of an FD building, called *eye*, with its six telescopes. *Right panel:* Schematic side view of a single FD telescope with its main components. Image adapted from Ref. [68]

Therefore, this technique has the advantage of having direct access to the energy of the primary particle, and to the shower maximum $X_{\max}$. However, the duty cycle of FD is about 15%, as the observation periods are limited to dark nights and the presence of the moon has the effect of contaminating the observed signal with additional light. In addition, as the atmosphere acts as a huge calorimeter, the presence of clouds, thunderstorms, aerosol, or dust affects the measurements. Hence, monitoring the atmospheric conditions is of great importance, and for this reason, to continuously monitor and study the atmospheric properties several instruments have been deployed at the Observatory [70].

Each fluorescence building, containing six telescopes, is called an *eye*, and it has a field of view of 30° in elevation and 180° in azimuth (30° in azimuth per telescope). Fig. 3.2 shows a schematic representation of an eye, together with a schematic representation of the fluorescence telescope. The air-fluorescence light enters through a UV-passing filter and is focused by a 10 m^2 spherical mirror into a camera of 20×22 Photo-Multiplier Tubes (PMTs) sensors (pixels).

The FD uses a hierarchical trigger chain to select real air-shower events. However, an FD-only detection suffers from a high uncertainty in the determination of geometrical properties of the shower, mainly because it is hard to determine the distance to the shower with a single-telescope measurement. For this reason, the PAO has been designed as a hybrid detector, with the SD providing essential information about the core location of the shower, and offering the possibility of cross-calibrating these

two-detector systems.

3.2.1 The HEAT enhancement

HEAT, with its three telescopes, is the low-energy extension of the fluorescence detector (FD) of the Observatory [71, 72]. As low-energy showers develop high in the atmosphere, usually $X_{\max}$ is not directly accessible, or in other words, outside the field of view of the standard FD telescopes. For this reason, though the optical system is the same as the standard FD, these telescopes are held in shelters, which in turn are mounted on top of a hydraulic mechanism capable of changing the elevation of the telescope optical axis. In their "upward" position, the field-of-view elevation range is $[30^\circ, 60^\circ]$, which in combination with the Coihueco field of view gives an elevation coverage from 0° to 60°.

Recently, a measurement of the CR energy spectrum has been performed in the $10^{16.5}$ - $10^{18.5}$ eV energy range using data from the standard FDs and HEAT [73]. As we shall see in Sec. 5.4, this energy range nicely overlaps with that of another low-energy enhancement of the Observatory, namely the 433 m SD Infill array. Although the energy calibration of this array could not be done against HEAT energy measurements, the energy spectrum measured by HEAT is taken as a reference to check the capabilities of this SD extension.

3.3 Surface Detector

The complementary technique used at the Observatory is the SD array [74]. Each detector in the array, every small black dot in Fig. 3.1, is a Water-Cherenkov Detector (WCD): a large polyethylene tank filled with purified water and coated with a reflective Tyvek-film liner on the inside. Fig. 3.3-*left* shows an actual picture of an SD station, and its schematic view is shown in Fig. 3.3-*right*, where the other components are indicated as well. The power is provided locally by a solar panel loading the batteries, and a Global Positioning System (GPS) receiver is used to establish a common time among the different stations.

As secondary particles generated in an EAS travel at nearly the speed of light in vacuum, their speed in water exceeds the speed of light in this medium ($\simeq 0.75c$ in water), giving rise to Cherenkov radiation inside the tank. The amount of emitted light is proportional to the length of the tracks drawn by the particles, and ultimately to the number of par-

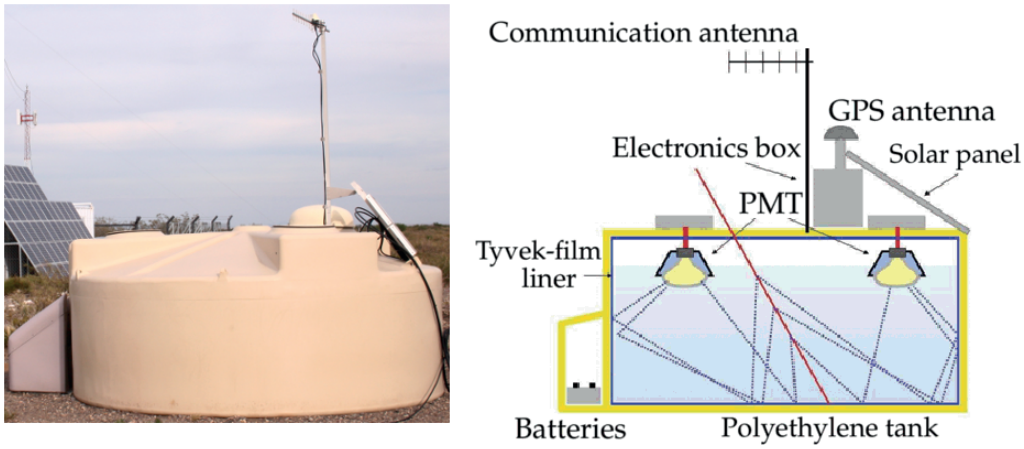

FIGURE 3.3: *Left panel:* A picture of a water-Cherenkov detector, an individual unit of the Observatory SD array. *Right panel:* Schematic cross section of the detector, with indicated components. Image adapted from Ref. [75]

ticles crossing the tank. This generated light is detected by three 20.3 cm diameter PMTs, equipped with a base having two outputs: the anode and the last dynode. These signals from the PMTs are filtered and digitized by Flash Analog to Digital Converters (FADCs), and read by a Programmable Logic Device (PLD) performing trigger decisions. Finally, when requested, these data are sent to the CDAS, which combines the information from the local stations.

Detector Calibration

The trigger condition for the local stations needs to be maintained at the same level for each station in the array. This condition is fulfilled if the station is able to locally set a trigger threshold in detector-independent units, which needs to be the same for all SD stations. For this reason the total signal is measured in units of the signal produced by a Vertical and Central Through-going (VCT) muon, conventionally called Vertical Equivalent Muon (VEM). Therefore, the aim of the calibration is to obtain the value of this reference unit, i.e. 1 VEM, in electronics units for each individual local station.

As the signal produced in the tanks depends on the station specifications (e.g. liner reflectivity and PMT gain, among others) the calibration must be performed by the local electronics. Moreover, the calibration cannot be dependent on the number of functioning PMTs, or on other differences

among stations.

It follows that the fundamental calibration information needed by the SD is the average charge, Q_{VEM}, as measured for a VCT muon. For this purpose, atmospheric muons at a rate of $\approx$ 2.5 kHz are used to precisely measure 1 VEM. However, the detector cannot discriminate vertical from inclined muons passing through it. Nevertheless, the distribution of produced light by isotropic atmospheric muons also generates a peak in the charge distribution [76] (cf. Fig. 3.4-*left*). Using a muon telescope providing the trigger in a reference tank ([77]), this peak has been measured at $Q_{\mathrm{VEM}}^{\mathrm{peak}} = 1.03 \pm 0.02$ VEM for each PMT, which is a measure of that portion of the total signal deposited close to the PMT. For the sum of the three PMTs the peak is at $\approx$ 1.09 VEM, and represents a measure of the total signal deposited in the tank.

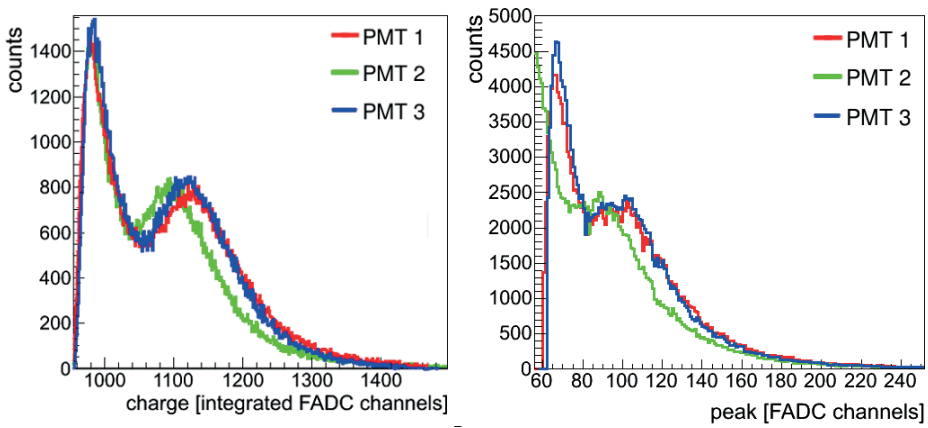

FIGURE 3.4: Example of calibration histograms. The peak at an abscissa value of 1000 is due to the very large number of low-energy particles present in the atmosphere. The other peak is caused by atmospheric muons. *Left panel:* Example of integrated charge histogram for the three PMTs of an SD station. *Right panel:* Example of pulse-height histogram for the same PMTs.

Next to the charge, a measure of the current from the PMTs is used to set the trigger levels. Therefore, a reference unit for this current needs to be defined as well. Similarly, this unit is the pulse height of light from VCT muons, I_{VEM}. Again, atmospheric muons are used to supply such a reference, as they produce a peak $I_{\mathrm{VEM}}^{\mathrm{peak}}$ in the pulse-height histogram that is used as the common reference unit for threshold levels (cf. Fig. 3.4-*right*). To maintain the proper trigger level, the conversion from channels to $I_{\mathrm{VEM}}^{\mathrm{peak}}$ has to be updated constantly.

The three main steps for the calibration to VEM units are:

(*a*) The end-to-end gains of each PMT are initially set to have $I_{\mathrm{VEM}}^{\mathrm{peak}}$ at 50 channels. This step makes sure that the signals recorded by the three PMTs are similar in amplitude (*gain matching*).

(*b*) A continuous online local calibration is performed to determine $I_{\mathrm{VEM}}^{\mathrm{peak}}$ (in channels) to adjust the electronics trigger level. This step is mainly intended to adjust for drifts occurring after the previous step.

(*c*) Determination of $Q_{\mathrm{VEM}}^{\mathrm{peak}}$ from the charge histograms. After converting the peak to 1 VEM, it is possible to convert the integrated signal of each PMT to VEM units.

Surface Detector Trigger

To discriminate air-shower events from single atmospheric muons, the triggering system of the SD consists of different trigger levels arranged in a hierarchical structure. At each level the conditions to be satisfied are more strict than the previous level, in order to minimize the presence of background events against high-energy showers. This hierarchy is organized as follow:

Single-station trigger → Array trigger → Physics trigger → Quality trigger

The **Single-station trigger** has two levels, the T1 evaluated by the PLD unit, where one of the following conditions needs to be satisfied by the station:

- a simple *threshold* trigger: the signal level has to be larger than the threshold at 1.75 VEM in all three PMTs.
- A *time-over-threshold* trigger (ToT), which is satisfied if at least 2 PMTs record a signal exceeding 0.2 VEM in a minimum of 13 FADC bins within a sliding window of 3 μs (120 bins). With a rate of $\approx$ 1.5 Hz, this condition selects small time-spread signals far from the shower core.

The second-level trigger is the T2, locally implemented in the software of the tank, is the logical OR between:

- a second-threshold trigger (TH), which is satisfied if the signal from

all 3 PMTs exceeds the set threshold of 3.2 VEM. With a rate of $\approx$ 20 Hz, this condition has been designed to select either high signals close to the shower core or pure muonic signals from horizontal showers.

- A T1 ToT. This trigger is automatically promoted to a T2 with no additional requirement.

The **Array trigger** (the T3) is formed by the CDAS. It has been designed to select real showers, searching for time coincidences among the signals in different local stations. Any station with a T2 sends the trigger-time information to the CDAS that searches one of the following conditions within the whole array:

- a 3-fold condition: it requires the coincidence, within a distance-dependent Δt, of three tanks holding a ToT condition.
- A 4-fold condition: it requires the coincidence, within a distance-dependent Δt, of four tanks passing any T2 provided that two tanks are found in the first two hexagonal crowns, and the other tank within four crowns (a crown being the set of stations at the same distance from the central one sending the trigger time).
- A further 3-fold condition requiring the coincidence of three aligned tanks passing any T2 condition.
- An external condition generated by the FD.

The **Physics trigger** (the T4) is the first trigger capable of discriminating real air showers from the background of random coincidences. For this purpose, the time delay among the candidate tanks has to be compatible with the speed of light, or in other words, "the difference in their start time has to be lower than the distance between them divided by the speed of light" (+200 ns). There are two types of T4, namely:

- *compact* 3ToT: it requires at least three stations with a ToT trigger in a triangular pattern (cf. Fig. 3.5).
- 4C1: it requires four tanks with a T2, with three stations in the first crown (cf. Fig. 3.6).

The **Quality trigger** (the T5) has been realized to select showers fully contained in the array, and hence excluding partially-sampled showers at the edges. This trigger requires that the station with the largest signal has the six neighbour stations regularly functioning at the triggering time, though not necessarily passing any trigger condition. This condition en-

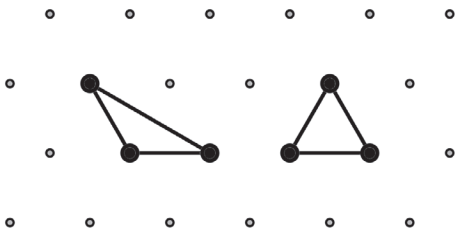

FIGURE 3.5: The two possible *compact* 3ToT configurations (with addition of all of the symmetry transformations of the triangular grid). Image taken from Ref. [78].

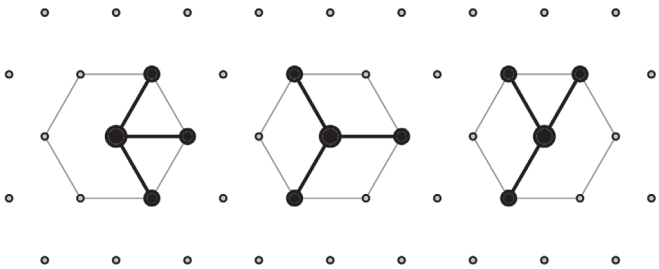

FIGURE 3.6: The three (minimal) 4C1 configurations (with addition of all of the symmetry transformations of the triangular grid). Image taken from Ref. [78].

sures a good and unbiased event reconstruction, and in addition, reduces the calculation of the exposure to the determination of the geometrical aperture and observational time.

3.3.1 The AMIGA enhancement

As briefly mentioned in Sec. 3.1, the AMIGA enhancement has been deployed to lower the energy threshold of the SD. It consists of a detector infill area of about 27 km^2 with 71 water-Cherenkov detectors on a 750 m grid. In Fig. 3.1 it is recognisable as the dense area on the left-hand side of the array, overlooked by Coihueco and HEAT. This arrangement reduces the energy threshold by a factor ten compared to the regular 1500 m array [79]. Part of the enhancement consists of underground scintillator detectors, buried close to the Infill stations to access directly the muonic component of air showers simultaneously measured with the SD.

Recently, the energy calibration of the Infill array has been carried out using a large dataset of coincident events with the fluorescence telescopes at Coihueco and HEAT [80]. Consequently, thanks to this enhancement, the CR energy spectrum measured at the Observatory has been extended by one order of magnitude in energy. In Sec. 5.3, the Infill measurement

of the CR energy has been taken as reference to perform the energy calibration of the 433 m SD Infill array.

3.4 Radio Detector: AERA

The AERA enhancement at the PAO has been proposed in March 2009 by a collaboration of research groups from France, Germany and the Netherlands. After the preliminary tests performed with small-size prototypes from 2007 to 2010, the deployment of AERA has begun in 2010 and currently it is taking data from March 2011. The full set-up consists of about 160 antennas distributed over an area of approximately 20 km^2.

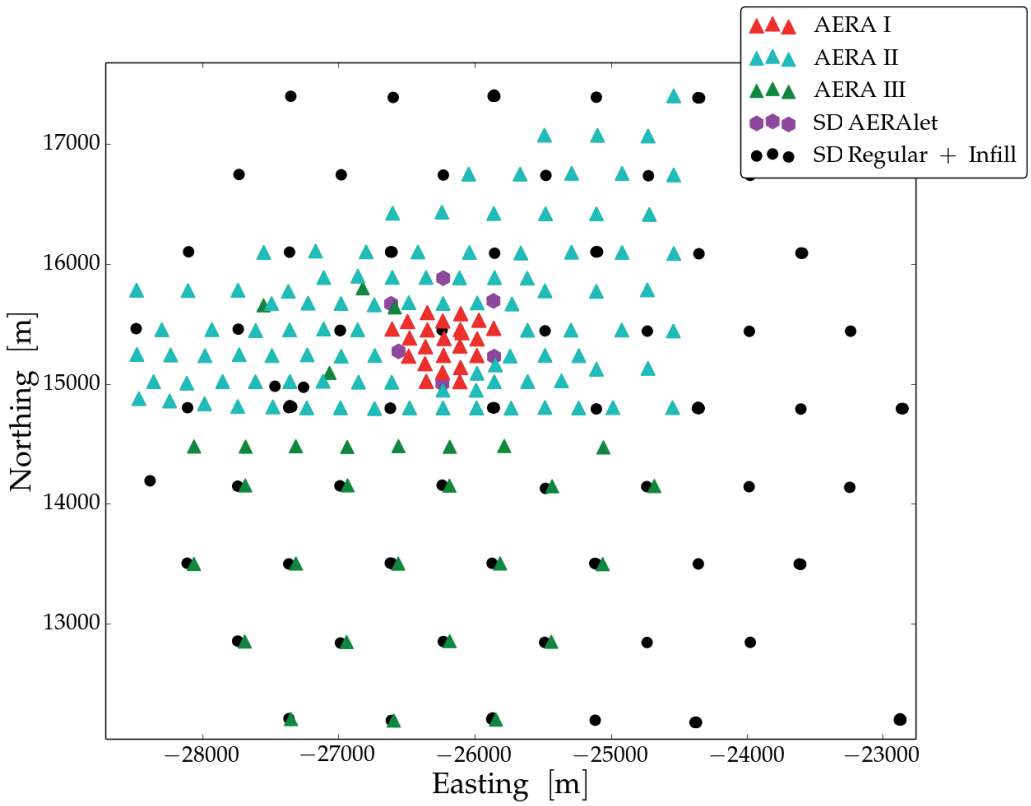

FIGURE 3.7: Layout of AERA with the additional AERAlet stations (in magenta), shown together with the SD Infill array. The origin of the coordinates is located at the center of the array (the *PampaAmarilla* Coordinate System).

Its deployment has gone through several phases, as shown in Fig. 3.9, as different antenna types and hardware versions have been tested [81]:

AERA I: the first stage of the installation, consists of the first 24 Radio-Detection Stations (RDSs) arranged on a triangular grid with a station-to-station spacing of 144 m, which form the core of AERA (the red triangles in Fig. 3.7). These stations are are equipped with Log-Periodic Dipole Antennas (LPDAs) (cf. Fig.3.8-*left*) optimized for radio frequencies in the 30–80 MHz band, and with low sensitivity towards the horizon. One of the antenna-polarization directions is aligned along the geomagnetic north-south (NS) direction (with a 0.6° accuracy), while the other points to the east-west (EW) direction perpendicular to the first. The signals from both polarizations are amplified by analog electronics that also suppress strong lines in the high-frequency band below 25 MHz and in the FM-broadcast band above 90 MHz. To convert the analog signal to a digital one, the stations are equipped with 12 bit digitizers running at a sampling frequency of 200 MHz or 180 MHz depending on the hardware type.

AERA II: the second stage of deployment consisted in the installation of 100 RDSs on a triangular grid. Part of these stations with a mutual distance of 250 m, and part at 375 m from each other. Though the frequency band is the same as the AERA I antennas, these stations are equipped with butterfly antennas [82], [81] (cf. Fig.3.8-*right*), the signal is digitized at 200 MHz or 180 MHz depending on the hardware type.

AERA III: has been the last stage of the deployment with the installation of 25 additional RDSs, of the same type of those of AERA II, but arranged on the SD Infill grid (the green triangles in Fig. 3.9).

As an engineering array, the AERA stations do not share all the same characteristics in terms of implemented triggers and communication systems. Most of the AERA I stations are connected via an optical glass fibre to the Central Radio Station (CRS), within the core of AERA, where one of the two Data AcQuisition systems (DAQs) is being run. The other stations communicate with the other DAQ, located at the FD Coiheuco, via a commercial wireless system using a regular Ethernet protocol.

Three triggering schemes have been implemented for the radio array, operating on different station clusters:

- *Self-triggers:* the triggering is performed on the radio pulse itself, based on a pulse identification algorithm in combination with a time-differences compatibility check between stations whose recorded pulse is larger than a certain threshold.

FIGURE 3.8: *Left panel:* An AERA I station. The log-periodic dipole antenna (LPDA) above the horizon, and the electronic box underneath the solar panels to the right. *Right panel:* An AERA II station equipped with the butterfly antenna, and the other visible antenna for communications. The solar panel is part of the electronics box hanged to the central pole.

- ▷ *Scintillator triggers:* some stations are equipped with scintillantors inside the electronics box, providing the trigger as particles cross them.
- ▷ *External triggers:* most of the stations are externally triggered by SD and/or FD.

In this thesis, we used only externally-triggered data by the 433 m SD Infill array, which is introduced in the next Section.

3.4.1 The AERAlet enhancement

This 433 m baseline SD Infill array was proposed as a low-energy enhancement of the PAO. It has been used to lower locally the energy threshold of the SD and to improve, at the AERA site, the accuracy on the determination of the core position of air showers.

The deployment of this 433 m array has started in November 2011, when the first three stations (Ids 97, 98, 99) have been installed around the Local Station (LS) Kathy Turner (Id 1764), which is already part of the 750 m Infill array of Auger. Although these first four stations have been collecting data since December 2011, at the end of August 2012 a new trigger algorithm has been implemented in the CDAS. This dedicated trigger has been called *AERA* (as the radio array, unfortunately), and it was designed

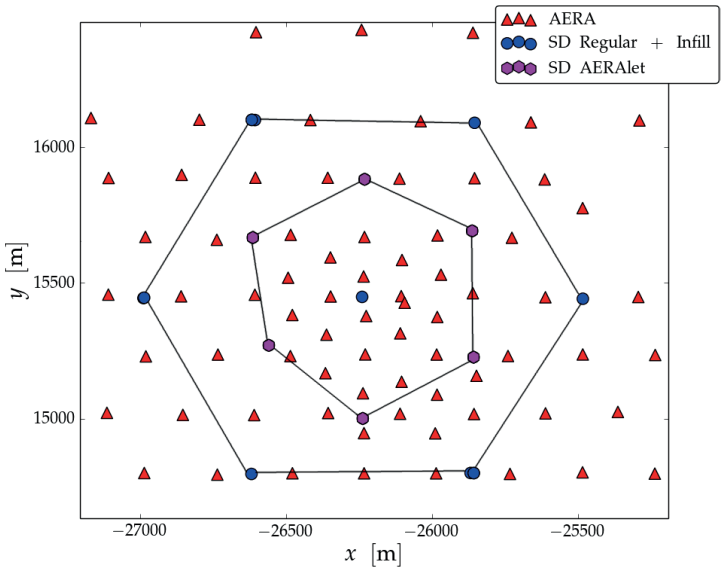

FIGURE 3.9: Layout of AERAlet, located at the center of AERA (red triangles). The origin of the coordinates is located at the center of the SD array (also known as the *PampaAmarilla* Coordinate System).

to allow the detection of air showers by means of the 433 m array only, i.e. it only triggers with these stations, and only requesting them. Indeed, such a narrow-spaced array has a very high rate, and the inclusion of other stations in the events will cause data-stream overload.

In January 2013 another three tanks (Ids 11, 12, 13) have been placed in the field, so the first hexagonal cell of this narrow array has been completed: six additional tanks around the central one, Kathy Turner (cf. Fig 3.9). These seven SD stations constitute the unique elementary cell of the 433 m infill array, which has been called AERAlet, being located at the site of AERA. Finally, the whole hexagon came into full operation only at mid May 2013, due to persistent communication problems.

In this thesis, air-shower data taken with this array have been used extensively to establish to what extent this facility of the Observatory may add valuable information at the low-energy regime around 10^{17} eV.

3.5 Conclusion

The Pierre Auger Observatory plays a major role in UHECR research. Thanks to its hybrid concept, which includes four different detection techniques, we are living the days where the Pierre Auger Collaboration

is going towards a complete characterization of EASs. This statement mostly relies on the possibility of performing independent measurements of a single CR event, by simultaneously accessing the different air-shower components. In addition, the upgrade of the SD array of the Observatory will give soon an unprecedented capability of discriminating the muonic and the electromagnetic signals.

Indeed, the presence of multiple air-shower-detection techniques at the Observatory will allow the possibility of observing *super-hybrid* events: UHECR events detected with all the available techniques. Evidently, such a unique opportunity represents an actual breakthrough in UHECR research.

"Logic is the foundation of the certainty of all the knowledge we acquire."

L. Euler

4

Reconstruction of Cosmic-Ray Events

The *reconstruction* of a CR event is the essential piece that connects the pure data-taking part with the high-level data analysis regarding the primary CR observables. Therefore, it is essential to give an overview on how the reconstructed shower parameters are inferred from the observations, as one mainly relies on the reconstruction software.

Therefore, this Chapter focuses on the *reconstruction* of CR events by means of the $\overline{\text{Off}}\underline{\text{line}}$ software framework [83], and for this reason it is mainly based on its own documentation. After introducing the software used by the Auger collaboration, the following Sections first focus on the event-reconstruction process based only on SD measurements used in Chapter 5. Second, this reconstruction process has been extended to allow the analysis of coincident SD and radio measurements used in Chapter 6.

4.1 The $\overline{\text{Off}}\underline{\text{line}}$ Framework

The $\overline{\text{Off}}\underline{\text{line}}$ framework has been designed as a *modular* software, fully configurable by the user by means of `XML` files. Together with the processing modules, which acts as the "workforce" part, the two other main components are the *event* data structure, and the *detector description* part.

In a very simplified way, the modules contain the algorithms able to read the detector/raw-event information, to process/combine this information, and to write the outcome back into the event data structure.

The modularity feature of Offline, which is the key of its flexibility, enables physicists to implement a large variety of applications to run over the raw data collected by the different detection techniques at the Observatory. Indeed, every *application* is mainly characterized by a single `XML` file consisting of a *sequence* of modules. Each module listed in the sequence file interfaces with the event data structure rather than with the other modules in the sequence, which then only communicate between them via this data structure. In this way, the user can simply adapt the sequence to its own necessity adding, removing, or reorganizing the modules. At the same time, each module of the module-sequence file, can be in turn configured by an `XML` file. The whole set of the configuration information, including the sequence file, is specified in a so-called *bootstrap* file, which is passed to the application at its execution.

Finally, the detector description consists of an interface, similar to the event one, to allow the modules to access the detector configuration and status, and all the monitoring databases mentioned in Chapter 3. This access is handled by a registry of *managers*, which does the real job of extracting the requested information from the specified data source. The desired managers, are as well specified in the bootstrap file, which then contains all the user specifications to run a particular application. Therefore, the use of bootstrap files largely simplifies the job of keeping track of the different configurations used, especially when running different applications on the same set of data.

In the Sections below I introduce in general terms the Offline applications used in this thesis, and in some details the essential algorithms used. Though user specifications will be specified in the relevant Chapter, where the reconstructed events are analysed, some of those are also given here, provided that this information does not interfere with the flow of the discussion.

4.2 SD-Only Reconstruction of Cosmic-Ray events

Here, the pipeline reported in Appendix A.1 is discussed for the reconstruction of CR events using only the information from the SD. The events reconstructed as specified below, are then used for the analysis in Chapter 5. Although there is some difference in the module configurations,

this pipeline is the standard one used by the collaboration to reconstruct SD-only raw data.

Once that the `EventFileReaderOG` and `EventCheckerOG` modules have respectively accessed and checked the raw-data file, the single-station calibration procedure, mentioned in Section 3.3, to convert the signal to VEMs is performed by the `SdCalibratorOG` module. In addition, this module flags the stations with zero-error code and without random triggers as *candidates*. Only these stations will be then considered by the subsequent modules. As the names suggest, the 'correction' modules, `TriggerTimeCorrection` and `SdStationPositionCorrection`, have been respectively used to correct for a small time offset in the PLD software and for the geographical position of some stations, which was erroneously reported.

The task of selecting physics events is committed by the `SdEventSelectorOG` module. As the Observatory includes SD arrays with different baselines (1500, 750, and 433 m), one needs to specify in the configuration file of this module, which array has to be considered. Indeed, for every event in the data file, it iterates over the candidate-station list to check whether the station is part or not part of the chosen array. In the latter case the station is flagged as *off-grid* and it will not be considered by the subsequent modules performing the actual reconstruction. Moreover, this module checks for the presence of so-called lighting stations (extremely high and oscillating signals), which are removed from the event, and of isolated stations, which are flagged as *accidental*. After checking the station T2 and array T3 conditions, the module calculates the T4 physics trigger, whose type (basically the configuration of the stations) is written on file according to the explanation given in Section 3.3. Subsequently, if enabled in the configuration, the (prior-to-reconstruction) T5 physics trigger is calculated and stored as well.

After the selection of stations, the main stages for the reconstruction of an air-shower properties are:

- Estimation of the shower arrival direction (axis) assuming a planar shower front moving at the speed of light, by the `SdPlaneFitOG` module.
- Determination of the shower-core position, and of the signal lateral dependence by the `LDFFinderKG` module.
- Refining the arrival-direction estimate by modelling the shower front as an expanding sphere, using the `LDFFinderKG` module.

The physics behind the algorithmic implementation is briefly discussed in the next two Sections.

Once the event has been successfully reconstructed, a further selection is implemented in the `SdPosteriorSelectorOG` module. The *posterior*-T5 condition, shown in Fig. 4.1, requires 5 working tanks (or a user-defined number) around the tank with the largest signal, independently from having a signal or not. In addition, it requires that the reconstructed core is enclosed in an equilateral triangle of functioning stations.

The last two modules, briefly discussed in Section 4.2.2 below, are needed to implement the plots for the `EventBrowser` and to add the reconstructed data to the event structure.

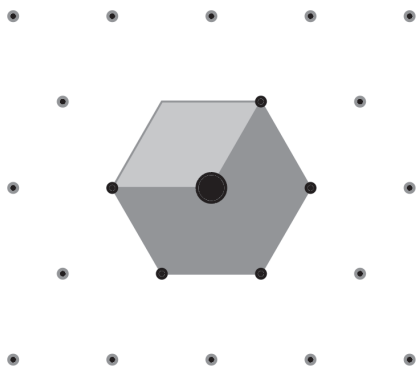

FIGURE 4.1: The posterior T5 condition. Image adapted from Ref. [78].

4.2.1 Reconstruction algorithms

Planar approximation

A first estimate of the CR arrival direction consists in approximating the shower-front to a plane moving at the speed of light, and in identifying the direction (θ, ϕ) with the shower direction. The time and spatial origins, from which all times and distances are measured, are respectively set as the weighted-signal bary-time and barycenter $c\tau_0, \boldsymbol{\xi}_0$, which is assumed to be the shower *impact point* on the ground. The shower axis is then the unitary vector $\hat{\boldsymbol{a}} \equiv (u, v, w)$ at the impact point, perpendicular to the shower plane and pointing towards the source of the particles.

For a planar shower front coming from the direction $\hat{a}$, the arrival time at a particular position $\xi(\tau)$ with respect to the barycenter, $\xi_0(\tau_0)$ are related

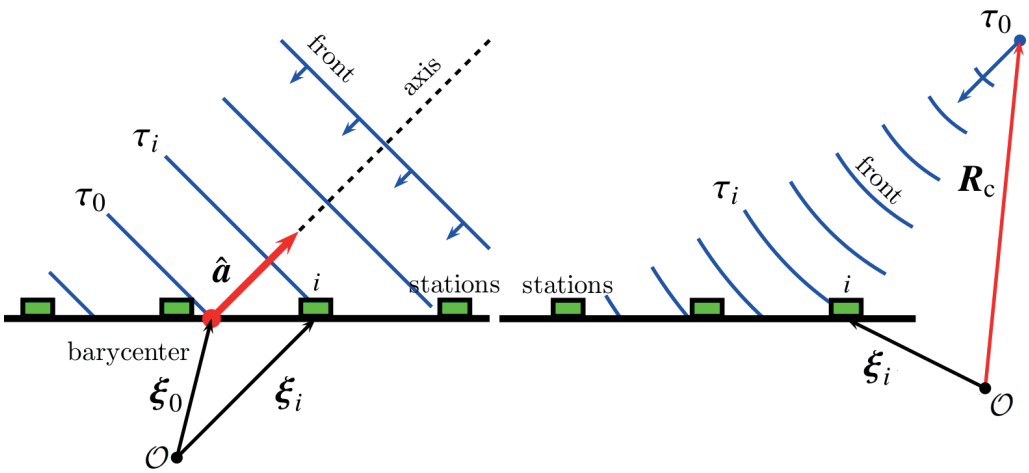

FIGURE 4.2: *Left panel:* Schematic illustration of the planar-front approximation used for a first estimate of the air-shower incoming direction. *Right panel:* Schematic illustration of the spherical-front approximation used to improve the shower-direction estimate. Image adapted from Ref. [84].

by

$$c(\tau - \tau_0) + \hat{\boldsymbol{a}} \cdot (\boldsymbol{\xi}(\tau) - \boldsymbol{\xi}_0) = 0. \tag{4.1}$$

Therefore, the time τ at which the shower plane passes through a generic point $\boldsymbol{\xi}$ can be inferred as

$$c\tau(\boldsymbol{\xi}) = c\tau_0 - \hat{\boldsymbol{a}} \cdot (\boldsymbol{\xi} - \boldsymbol{\xi}_0). \tag{4.2}$$

Specifying the generic points $(\tau, \boldsymbol{\xi})$ as the station positions $\boldsymbol{\xi}_i \equiv (\xi_i, \upsilon_i, \zeta_i)$ and the signal start times τ_i (cf. Fig. 4.2-*left*), which are measured, the unknown variables in the Equations above are only the components of the shower axis $\hat{\boldsymbol{a}}$. It follows that to obtain the shower-axis components, the shower front model can be fitted by minimizing the sum of the (squared) differences between the measured start times τ_i and the expected times (by Eq. (4.2)), weighted by the start-time variances $\sigma^2_{\tau_i}$ [85]

$$\begin{aligned} \sum_{i=1}^{N_{\text{stations}}} \frac{\{\tau_i - \tau(\boldsymbol{\xi}_i)\}^2}{\sigma^2_{\tau_i}} &= \sum_{i=1}^{N_{\text{stations}}} \frac{\{c\tau_i - c\tau_0 + \hat{\boldsymbol{a}} \cdot \boldsymbol{\xi}'_i)\}^2}{c^2\sigma^2_{\tau_i}} \\ &= \sum_{i=1}^{N_{\text{stations}}} \frac{\{c\tau_i - c\tau_0 + u\xi'_i + v\upsilon'_i + w\zeta'_i)\}^2}{c^2\sigma^2_{\tau_i}}, \end{aligned} \tag{4.3}$$

where $\boldsymbol{\xi}'_i = \boldsymbol{\xi}_i - \boldsymbol{\xi}_0$. The minimization above is constrained by the condition $w = \sqrt{1 - u^2 - v^2}$, since $\hat{\boldsymbol{a}}$ is unitary, which yields a non-linear

equation. However, an (approximate) analytical solution can be found assuming the stations all being close enough to a certain plane (the tangential plane to the ground containing the barycentre), for which the ζ-component can be neglected (as $\zeta_i \ll \xi_i, \upsilon_i$).

Determining the LDF

Once the shower axis vector has been estimated, it is possible to determine the lateral dependence of the measured signals, as the model depends on the distance r of the station to the shower axis (i.e. the distance in the shower plane). The chosen LDF is the `NKGBetaOnly`, a modified version of the Nishimura-Kamata-Greisen (NKG)-type function [86],

$$S(r) = S(r_{\text{opt}}) \cdot f_{LDF}(r) = S(r_{\text{opt}}) \cdot \left(\frac{r}{r_{\text{opt}}}\right)^{\beta} \left(\frac{r + r_{\text{opt}}}{r_{\text{scale}} + r_{\text{opt}}}\right)^{\beta}, \tag{4.4}$$

where, by construction, $f_{LDF}(r_{\text{opt}}) = 1$, and β governing the signal drop with increasing distance, or in other words the LDF *slope*. The optimal distance r_{opt} is defined as the distance on the shower plane where the slope (β) fluctuations minimally affects the signal model. In other words, the model-provided signal at this distance is maximally reliable, as it is the point where changes in the slope (induced by shower-to-shower fluctuations and statistical fluctuations in the number of particles) have the minimum influence on the LDF. This parameter depends on the array geometry, in particular on the mutual distance between the stations. In Section 5.2.3, this parameter has been fixed to 270 m for the AERAlet array, while it is 450 m for the Infill array [87], and 1000 m for the Regular array [88]. The scale parameter r_{scale} plays a role only at larger distances from the shower axis, and in addition it is strongly correlated to β, hence it has been kept fixed to its hard-coded value of 700 m.

The signal uncertainties, originating from sampling and shower-to-shower fluctuations, are described as a Poisson-like term, normalized by a zenith-angle dependent factor $\propto \sec\theta$, as the signal uncertainty depends on the ratio of the electromagnetic and the muonic components. Indeed, for inclined showers the relative muonic-component contribution to the signal increases, together with the average track length in the tank [89].

If requested by the user, the slope β may be left as free parameter as well. However, if the number of stations participating in the event is less than 5, the slope cannot be fitted with enough precision[1]. In this latter case,

[1]Other conditions need to be satisfied in order to properly fit β. These conditions are listed in Section 5.2.1,

the slope is fixed by the following model

$$\beta(\log_{10} S_{r_{\rm opt}}, \theta) = a + b \log_{10} S_{r_{\rm opt}} + (c + d \log_{10} S_{r_{\rm opt}}) \sec\theta + (e + f \log_{10}) \sec^2\theta \tag{4.5}$$

where θ is the shower zenith angle, and $S_{r_{\rm opt}} \equiv S(r_{\rm opt})$. The six parameters above depend as well on the array geometry, and their determination for the AERAlet array will be discussed in Section 5.2.3.

The aim of the `LDFFinder` module is to fit the signal model in Eq. (4.4) to the measured signals and distances (S_i, r_i) to determine $S_{r_{\rm opt}}$, needed for the energy estimation, and the shower-core position, needed for the curvature fit. The LDF fit is performed in Offline via a maximum-likelihood method, which allows to properly take into account zero-signal and saturated stations, and the signal fluctuations model.

The process of fitting the LDF is performed in several steps:

1. the geometry is fixed by the `SdPlaneFit` module, and $S_{r_{\rm opt}}$ is approximated by the signal of the closest station to $r_{\rm opt}$.
2. The slope is initially fixed by $\beta_{\rm init}(\theta) = 0.9 \sec\theta - 3.3$ [90].
3. The fitting of $S_{r_{\rm opt}}$ and shower-core location is performed. The shower-core z component is assumed to zero, and x, and y are allowed to float on the tangent plane containing the barycentre.
4. The complete maximum-likelihood fit is performed.

After that, the `LDFFinder` module uses the determined values of $S_{r_{\rm opt}}$ and core location to estimate the energy, and to refine the direction reconstruction by considering the curvature of the shower front.

As we shall see in Section 5.3.1, the energy estimate uses the value of $S_{r_{\rm opt}}$, though this parameter implicitly depends on the zenith angle because of the atmospheric attenuation of the shower particles. This dependency can be disentangled by means of the CIC method [91]. Assuming that the high-energy cosmic-ray flux is isotropic, $S_{r_{\rm opt}}$ is converted into a reference signal size $S_{\theta_{\rm ref}}$

$$S_{\theta_{\rm ref}} = \frac{S_{r_{\rm opt}}}{p(\tilde{\theta})} \qquad \text{with} \qquad \tilde{\theta} := \cos^2\theta - \cos^2\theta_{\rm ref}. \tag{4.6}$$

Again, the value of $\theta_{\rm ref}$ depends on the trigger-threshold energy of the array (and thus on the array geometry), and $p(\tilde{\theta})$ is the third-order poly-

as part of the event selection.

nomial determined by the CIC method with $p(0) = 1$. Thus, the energy estimate is given by:

$$E_{\mathrm{SD}} = A \cdot S^{B}_{\theta_{\mathrm{ref}}} \quad [\mathrm{EeV}], \tag{4.7}$$

with the calibration coefficients A, and B determined by the correlation between the estimator $S_{\theta_{\mathrm{ref}}}$ and the reference energy measured by FD, for events coincident between SD and FD.

The estimate of the incoming direction can be improved by approximating the curvature of the shower front with a spherical surface expanding at the speed of light, starting at time t_0 and at the point $\boldsymbol{R}_{\mathrm{c}}$. If $\boldsymbol{\xi}_i$ are the station positions, then for each station the time τ_i satisfies

$$c\,(\tau_i - \tau_0) = |\boldsymbol{R}_{\mathrm{c}} - \boldsymbol{\xi}_i|. \tag{4.8}$$

As we are interested in the shower axis $\hat{\boldsymbol{a}}$, one can parametrise the apparent origin of the shower as $\boldsymbol{R}_{\mathrm{c}} = \mathfrak{c} + R_{\mathrm{c}}\,\hat{\boldsymbol{a}}$, with R_{c} the radius of curvature as measured at the impact point $\mathfrak{c}$ (of which the barycenter $\boldsymbol{\xi}_0$ is a rough estimate). The minimization of

$$\sum_{i=1}^{N_{\mathrm{stations}}} \frac{\{c\,(\tau_i - \tau_0) - |R_{\mathrm{c}}\,\hat{\boldsymbol{a}} - \boldsymbol{\xi}'_i|\}^2}{c^2\sigma^2_{\tau_i}}, \tag{4.9}$$

with $\boldsymbol{\xi}'_i = \boldsymbol{\xi}_i - \mathfrak{c}$, can be achieved approximately in an analytical way or with a numerical non-linear optimization, to determine the radius of curvature R_{c} and accurate shower-axis components.

4.2.2 Data Storage and Visualisation

All the reconstructed information, together with the raw information, are finally written to Advanced Data Summary Tree (ADST) files by the `RecDataWriter` module. This data format is based on ROOT libraries [92], and it has been used for the main analysis in this thesis. Thanks to the `SdRecPlotter` module, the reconstructed events are easily visualized by means of the `EventBrowser` program, which comes along with the Offline framework. Using this browser, one can debug and understand the air-shower reconstruction output.

As an example, the main frames of the `EventBrowser` for a reconstructed event are shown in Fig. 4.3.

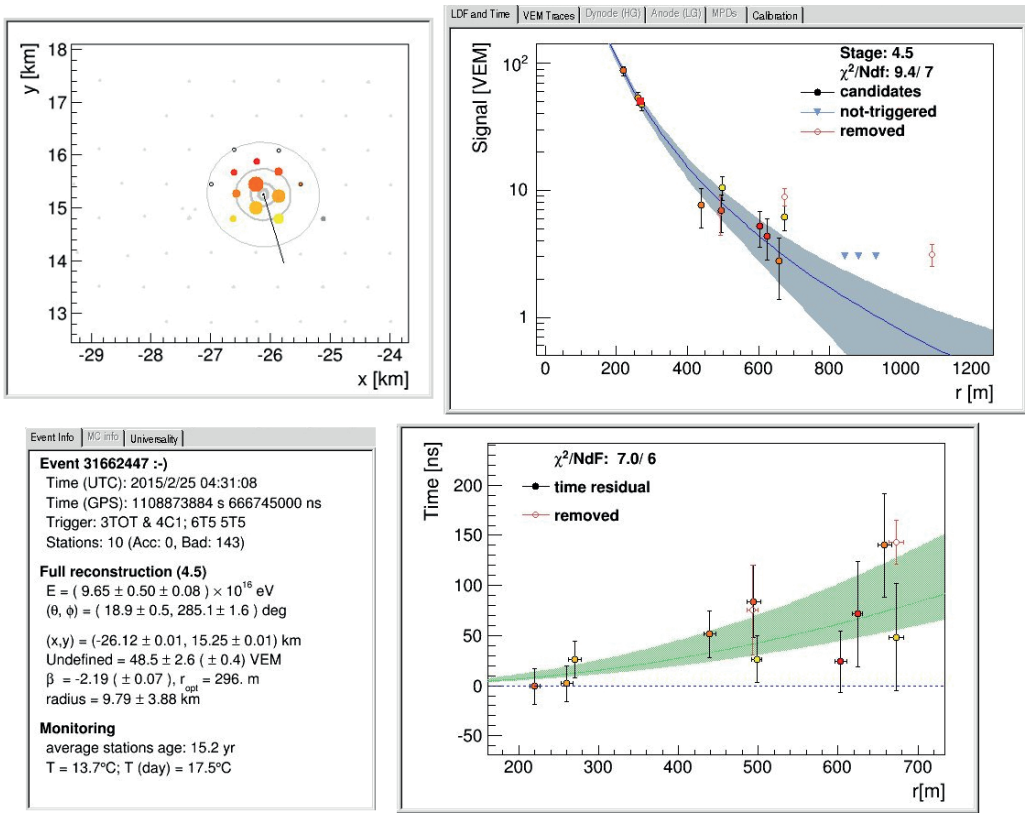

FIGURE 4.3: Main frames in the `EventBrowser` for a reconstructed CR event by means of SD data. The top-left frame shows a view of the array (the used array for this example is the unique hexagon of the 433 m SD Infill in combination with the first crown of the 750 m SD Infill array. See Section 5.3.2): candidate stations with signal are indicated by the filled circles, where the color is corresponding to the delay to the plane front arrival time, and the size to the signal magnitude. The top-right frame shows the fitted LDF, with the red square being $S_{r_{opt}}$, whilst the bottom-right the time residuals with respect to the spherical-front model. Finally, the bottom-left frame summarises the main CR-event information.

4.3 Radio-SD Reconstruction

The modularity feature of the Offline framework makes it an extremely versatile tool necessary to provide a large variety of applications required to analyse hybrid data. The following pipeline has been used to reconstruct coincident events measured by surface and radio detectors, on which the analysis in Chapter 6 is based. As most of the modules contains user-configurable parameters, the description of the pipeline reported in Appendix A.2 is given accordingly to the actual used configuration.

The first part is a typical SD-only reconstruction pipeline mainly used to determine the air-shower geometrical properties relevant for the radio modules. The primary aim of the radio part is the reconstruction of the time-dependent electric field vector removing all of the detector characteristics except for the limitation to a finite observing frequency window. For this purpose, at each antenna position the response of the antenna channel to the measured data needs to applied.

As for the SD-only reconstruction case, a first-level event selection is performed by the `RdEventSelector` module, to check whether enough antennas are participating in the event. This module allows one to select only externally-AERAlet-triggered radio events in the data stream, and to specify the minimum number of stations to be present in the event, which has been fixed to 1.

Once the data is read-in, the `RdEventInitializer` is called to set the coordinate origin at the core position estimated by SD, and to set the windows for signal and noise searches to be used by all later modules, relative to the SD shower-core time. The signal-search window has a width of 250 ns, while the noise region N is 3 μs wide, far off the signal region.

Afterwards, channel-wise ADC counts are converted to voltages by the `RdChannelADCToVoltageConverter` module, and a possible DC offset of the sampled trace is then removed by the `RdChannelPedestalRemover` module. Thereafter, the backwards detector response is calculated with the `RdChannelResponseIncorporator` taking the attenuation, amplification and phase delays due to the analogue components (i.e. cables, filters and amplifiers) into account. After these low-level detector effects have been removed the internal data structures contain the voltages at the foot-point of the antenna. Subsequently, the `RdChannelBandstopFilter` applies a bandstop filter to suppress narrow-band transmitters and other sources of Radio-Frequency Interference (RFI). This module sets the data of a specific frequency bin to zero if its amplitude is larger than (median) $+3\times$ (standard deviation), with median and standard deviation of the time samples in the selected trace.

A fundamental step for the reconstruction is the `RdAntennaChannelToStationConverter` module, which converts the channel voltages to an electric field using the antenna pattern and the arrival direction (from SD) as input [93]. As a result, the electric field E_j is calculated for every spatial dimension (EW, NS, and vertical), and for each sample number j in the time trace. Once the information is converted at the station level, the pulse parameters on the station level are determined and filled into the reconstructed data structure by the `RdStationSignalRecon-`

`structor` module. This module defines the station signal S as the amplitude of the signal in the bin with the highest amplitude, and the noise N as the Root Mean Square (RMS) in the noise window. In addition, it applies a signal-to-noise cut ($S/N > 10$) to flag stations as candidates, i.e. stations that likely have recorded a genuine radio pulse, though a further check is performed by the `RdStationQualityAssessor` module, which calculates the maximum amplitude outside the signal region.

4.3.1 Polarization Information

Finally, the `RdPolarizationReconstructor` module extracts the electric field polarization information by reconstructing the Stokes parameters I, Q, U, and V defined in Eq. (2.4) at the station level. For this purpose, the complex representation $\mathcal{E}_j$ of the electric field $\mathbf{E}_j$ is computed by

$$\mathcal{E}_j = \mathbf{E}_j + \mathrm{i}\tilde{\mathbf{E}}_j, \tag{4.10}$$

where $\tilde{\mathbf{E}}_j \equiv \mathcal{H}(\mathbf{E}_j)$ is the Hilbert transform [94] of the electric field. Before the Stokes parameters can be computed from the trace, one needs to transform the field in the reference frame where the vertical component is along the propagation vector of the wave (the Poynting vector), approximated by the arrival direction of the shower $\hat{\boldsymbol{z}} \equiv \hat{\mathbf{v}}$. In fact, as mentioned in Section 2.2.1, the radio emission is beamed into the direction of the shower axis as the source particles move in this direction at approximatively the speed of light.

Therefore, the plane perpendicular to the Poynting vector is the plane where the electric field oscillates, and on this plane one can choose as basis vectors

$$\hat{\boldsymbol{x}} = \frac{\hat{\mathbf{v}} \times \mathbf{B}_\oplus}{|\hat{\mathbf{v}} \times \mathbf{B}_\oplus|}, \text{ and } \hat{\boldsymbol{y}} = \hat{\mathbf{v}} \times \hat{\boldsymbol{x}}. \tag{4.11}$$

Having identified the propagation direction of the emitted radio wave with the shower axis, it follows that the plane spanned by the basis vectors above is exactly the shower plane (cf. Fig. 4.4). Once the traces are transformed in this frame, following Eqs. 2.4, the Stokes parameters are

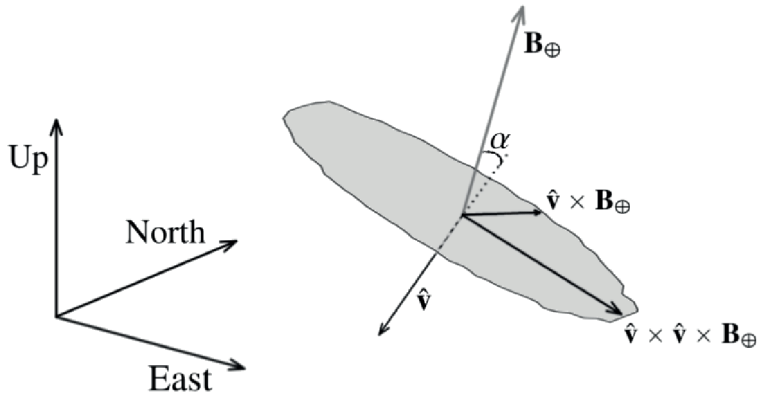

FIGURE 4.4: The chosen reference frame with the vertical axis along the direction of the incoming shower, indicated as $\hat{\mathbf{v}}$, and the $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ axis as defined in Eqs. (4.11). The (East, North, Up) is the Pampa Amarilla reference frame, where the antenna positions are usually specified. The direction of the geomagnetic field vector is indicated by $\mathbf{B}_\oplus$, and the angle between the shower axis and the geomagnetic field direction is denoted as α.

computed by

$$
\begin{aligned}
I &= \frac{1}{M}\sum_{j=1}^{M}\left(E_{j,x}^2 + \tilde{E}_{j,x}^2 + E_{j,y}^2 + \tilde{E}_{j,y}^2\right)\\
Q &= \frac{1}{M}\sum_{j=1}^{M}\left(E_{j,x}^2 + \tilde{E}_{j,x}^2 - E_{j,y}^2 - \tilde{E}_{j,y}^2\right)\\
U &= \frac{2}{M}\sum_{j=1}^{M}\left(E_{j,x}^2 E_{j,y}^2 + \tilde{E}_{j,x}^2 \tilde{E}_{j,y}^2\right)\\
V &= \frac{2}{M}\sum_{j=1}^{M}\left(\tilde{E}_{j,x}^2 E_{j,y}^2 - E_{j,x}^2 \tilde{E}_{j,y}^2\right),
\end{aligned}
\tag{4.12}
$$

where $M = 50$ samples for the signal window centred around the pulse maximum. These values are then corrected by subtracting the polarization of the noise (computed similarly as above but for the noise region [95]), and the parameter uncertainties are computed as in [54].

Finally, at each radio station the angle between the semi-major axis of the polarization ellipse and the $\hat{\boldsymbol{x}}$ axis, or simply the polarization angle ψ,

is calculated from U and Q as

$$\psi = \frac{1}{2} \arctan\left(\frac{U}{Q}\right), \tag{4.13}$$

where it is important to take care of the sign of U and Q separately, to identify the correct quadrant. Accordingly, the uncertainty σ_ψ on the polarization angle is obtained by

$$\sigma_\psi = \sqrt{\frac{\sigma_Q^2 U^2 + \sigma_U^2 Q^2}{4(U^2+Q^2)^2}}. \tag{4.14}$$

Together with the polarization angle ψ, each antenna position is projected on the shower plane, where its position can be represented in polar coordinates (r, ϕ). Here, r is the distance of the antenna in the shower plane (i.e. the magnitude of the planar vector $\mathbf{r}$ from the shower core to the projected antenna position), and ϕ the *observer angle* between this vector and the $\hat{\boldsymbol{x}}$ axis. Therefore, the observer angle is determined by

$$\phi = \arccos\left(\frac{\hat{\boldsymbol{x}} \cdot \mathbf{r}}{||\hat{\boldsymbol{x}}||\,||\mathbf{r}||}\right). \tag{4.15}$$

Though this formula does not distinguish between two angles having the same cosine, looking at the sign of the projection with the $\hat{\boldsymbol{y}}$ axis, $\mathbf{r} \cdot \hat{\boldsymbol{y}}$, one can tell whether $\phi \in (0, \pi]$ $(\mathbf{r} \cdot \hat{\boldsymbol{y}} > 0)$ or $\phi \in (\pi, 2\pi]$ $(\mathbf{r} \cdot \hat{\boldsymbol{y}} < 0)$. To estimate the uncertainty σ_ϕ of the observer angle, mainly due to the core-position uncertainty, a Monte Carlo (MC) simulation has been used, as the `RdPolarizationReconstructor` module does not provide this uncertainty. More details are given in Section 6.1.2.

As already mentioned in Section 4.2, all reconstructed quantities mentioned in the present Section are written by the `RecDataWriter` module to ADST files, together with all the other event- and shower-related quantities.

4.4 Summary

In this Chapter, we presented the necessary steps to convert the raw information at the detector level into the information concerning the primary CR, exploting the EAS properties described in Chapter 2.

This process is performed by means of $\overline{\text{Off}}\underline{\text{line}}$, the official software for

reconstructing CR events of the Pierre Auger Observatory. The event reconstruction is a necessary step for any high-level analysis that involves CR observables, such as the primary energy, and incoming direction. Therefore, it is extremely important to validate its performance as new functionalities are added to the framework. For this purpose the `EventBrowser` has proved to be a very effective tool, as it gives the possibility of checking the information of the reconstructed events stored in the ADST files.

All the analysis in the present thesis are based on the these files, by means a dedicated software developed in Python.

" Science exhibits itself as a circle returning upon itself, the end being wound back into the beginning, the simple ground, by the mediation; this circle is moreover a circle of circles, for each individual member as ensouled by the method is reflected into itself, so that in returning into the beginning it is at the same time the beginning of a new member. Links of this chain are the individual sciences [of logic, nature and spirit], each of which has an antecedent and a successor – or, expressed more accurately, has only the antecedent and indicates its successor in its conclusion."

G. W. F. Hegel

5

AERAlet Measurements of Cosmic Rays

As described in Section 3.4.1, the 433 m Surface Detector (SD) Infill array consists of six additional stations around the Local Station (LS) Kathy Turner (already part of the 750 m SD Infill array) in the middle of the Auger Engineering Radio Array (AERA), from which the short name AERAlet. As in the Central Data Acquisition System (CDAS) a dedicated central-trigger algorithm has been implemented for AERAlet, once the array was fully deployed, it has become possible to reconstruct cosmic-ray events detected by this dense array.

For this purpose, it has first been necessary to adjust those parameters, used in the Lateral Distribution Function (LDF)-fitting procedure that depend on the array geometry, i.e. the *optimal distance* r_{opt} and the *slope* β of the LDF (see Section 4.2).

After this step (Section 5.2), the Constant Intensity Cut (CIC) method has been applied to the newly reconstructed events, to allow for an estimation of the energy-calibration parameters for AERAlet (Section 5.3). Once a reliable estimation of the energy has been achieved, and the energy-calibration parameters inserted into Offline, a set of two years of data has been processed to obtain a low-energy extension of the flux spectrum (Section 5.4).

5.1 AERAlet Monitoring and Performances

As each SD station communicates its T2 trigger to the CDAS, which combines these to form higher-level triggers, the number of T2 triggers is stored every second in the so-called T2 files. These files are used to monitor the activity of each SD station second by second. As the expected T2 rate is $\approx$ 20 Hz (see Section 3.3), from this number the station *uptime* is easily calculated. In addition, this information is used to estimate the uptime for each hexagonal cell, which in turn allows the calculation of the instantaneous aperture of the SD (see Section 5.4.1). As an example, Fig. 5.1 summarizes the daily T2 uptime for each AERAlet station during the month of August 2013, with the red-horizontal line indicating an uptime of 24 hrs (i.e. an all-day fully-operational detector). Then, given the correlation among the station uptimes (visible in this Figure), the (daily) elementary-cell uptime is approximated by the minimum uptime among the stations being part of the considered cell (so for example on the 20^{th}, the uptime of AERAlet has been $\sim$ 10 hrs).

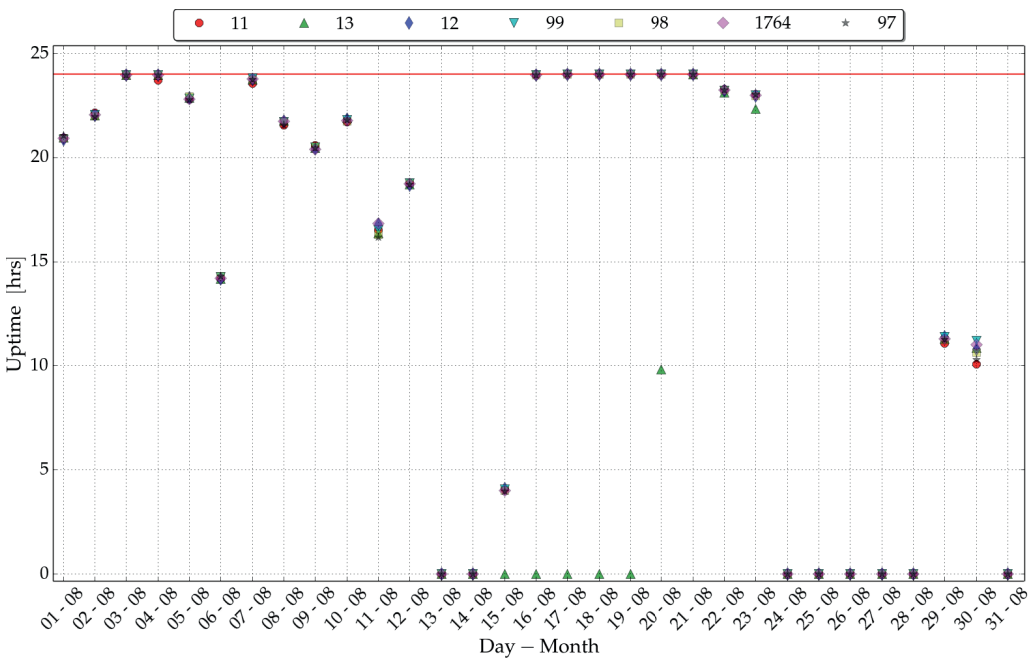

FIGURE 5.1: Daily T2 uptime of the seven AERAlet stations during the month of August 2013.

During the whole 2013, the T2 uptime of the AERAlet hexagon has shown some instabilities and only at the end of the year started to be

stable at $\approx$ 24 hrs. To give an idea of the performances for the AERA trigger, Fig. 5.2 shows the coincidence rate, in events per second, against the station multiplicity. The values in the plot have been calculated using *AERA* triggers, which occurred in August 2014 (the month with the most stable T2 uptime), for which the chance of recording random events is very low. Therefore, under the condition of a smoothly working detector, one approximately gets about 388 three-, 140 four-, 65 five-, 33 six-, and 17 seven-fold events per day.

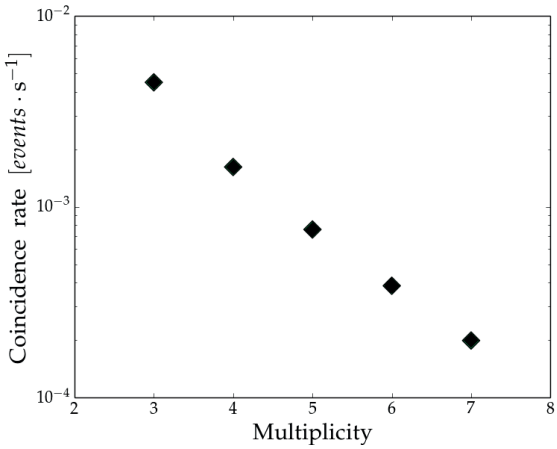

FIGURE 5.2: Coincidence rate for AERAlet as a function of the station multiplicity, extracted from one month (August 2014) of data using $\approx$ 24 hrs daily-stable T2 uptime per station.

5.2 Tuning the LDF Parameters

The present section focuses on the estimation of r_{opt} and on the determination of the LDF-slope parametrisation, which depends on the spacing between the stations of the array. In particular, along the reconstruction pipeline discussed in Section 4.2, the `SdEventSelector` and `LDFFinderKG` modules are the most affected ones that need to be configured accordingly to the actual spacing. Indeed, the first module is responsible for selecting the correct stations, and the second module performs the actual fitting of the LDF, and in turn of $S_{r_{\text{opt}}}$, which finally is needed to estimate the energy of the primary CR initiating the observed air shower.

Because of the flexibility of the `XML` configuration of $\overline{\text{Off}}\underline{\text{line}}$, it has been possible to leave the slope β as a free parameter to be fitted for those

events with a number of stations $\geqslant 5$. In the next Sections, we are going to see how, using these events, one can infer the optimal distance r_{opt}, and the parameters for the β model.

5.2.1 Event Selection

All CR events used in the forthcoming analysis, were detected by making use of the *AERA* trigger algorithm described in Section 3.4.1. As mentioned in that Section, the AERAlet hexagon has entered into operation at the end of January 2013; nevertheless a persistent communication problem with station 97 heavily reduced the amount of data. Moreover, another communication crisis, in April and May 2013, reduced the effective data-taking period as well. As a consequence, the T2 uptime was not constant during the considered data-taking period, i.e. from February 2013 to February 2015. Therefore, the collected dataset contains far fewer events than expected based on the (maximal) T2 trigger rate (about 50% less).

The selection of the CR events in which we are interested has been performed in two steps further explained below. First, to reduce the computational time, at the raw level only data written by the *AERA* trigger algorithm are selected. Second, a selection of physics CR events is performed in $\overline{\text{Off}}\underline{\text{line}}$, by means of the `SdEventSelector` module, before attempting the reconstruction.

As a matter of fact, a single-day `IoAuger`-type data file includes all types of events detected at any point in the whole Observatory. Therefore, a specific CDAS application has been designed to write raw files containing only *AERA*-triggered events recorded in the data stream. This pre-selection has been essential for the aim of reconstructing the whole dataset in a reasonable amount of time, and to be able to perform the reconstruction many times (e.g. after varying a single parameter in the $\overline{\text{Off}}\underline{\text{line}}$ configuration).

The second step of selection, i.e. selecting those events that can be actually reconstructed, is the standard procedure for finding physics events performed in $\overline{\text{Off}}\underline{\text{line}}$ by the `SdEventSelector` module (see Section 4.2). The configuration file used for this module is reported in Appendix A.3. A preliminary step consisted in introducing a third `<GridType>` for the `SDetector` class. The specified distances in the XML configuration have been deduced by interpolating the values of the Regular and Infill arrays. Therefore, every selection and trigger-identification algorithm in

this module will work as usual, once the correct settings have been specified.

Finally, the dataset of two years of data resulted in 151758 CR events, which still need to undergo the quality cuts. These cuts are mainly intended to exclude those showers falling at the edge of the AERAlet hexagon, which will unavoidably bias the analysis if taken into account, and to only include those events with fitted slope. In fact, in this case, as only part of the shower front will be sampled, the resulting LDF would represent only part of the actual LDF, and consequently the particle content would be underestimated and so the primary energy.

5.2.2 Event Reconstruction

The reconstruction of physical air-shower events detected by AERAlet has been performed with the Offline version v3r0p0, and the used reconstruction pipeline has been described in Section 4.2. As mentioned above, the sensitive parameters to the array geometry, r_{opt}, $S_{r_{\text{opt}}}$, and the slope-model parameters, enter in the functional form of the LDF, specified in Eq. (4.4)

Though the shower size $S_{r_{\text{opt}}}$ is fitted event-by-event by the `LDFFinderKG` module, the final value is returned as $\hat{S}(r_{\text{opt}})$ (i.e. as the model value at $r = r_{\text{opt}}$), where $\hat{S}$ is the signal model of the single event, after all stages of the LDF-fitting have been performed. Hence, the optimal distance needs to be specified *a priori* in the module configuration file through the `<ldfReferenceDistance>` tag (cf. Appendix A.4). To distinguish the prior choice from the optimal distance to be determined, we shall refer to the configured reference distance as D_{ref}. For this reason, a reconstruction of the dataset has been performed for each $D_{\text{ref}} \in \{200, 250, 300, 350, 400\}$, and the suitable distance r_{opt} for AERAlet has been determined analysing the reconstructed events (see Section 5.2.3).

As already mentioned, we wish to reconstruct events with β as a free parameter, which has meant setting the `<minNumberRelaxBeta>` configuration parameter to 5 (see A.4). This tag indicates the minimum number of stations required to check if a reliable fit of β can be attempted. Such a fit will be performed if the stations participating in the event fulfil the following additional requirements, where again the distances are set accordingly to the array spacing:

- 2 stations in the range $[125\text{ m}, 450\text{ m}]$ with minimal distance between

any pair $\geqslant$ 170 m **OR**
- 3 stations in the range [125 m, 450 m] with minimal distance between any pair $\geqslant$ 130 m **OR**
- 4 stations in the range [125 m, 450 m] with minimal distance between any pair $\geqslant$ 100 m.

The requirement of having a fitted LDF-slope is, among others, one of the quality cuts used for the following analysis.

5.2.3 Data Analysis

For the following analysis the applied quality cuts to the events are listed below:

i. Zenith-angle cut: $\theta \leqslant 55^{\circ}$.
ii. Events with no saturated stations.
iii. Events that fulfil the posterior T5 condition.
iv. Fully-reconstructed events with fitted LDF-slope β.

The amount of events passing these cuts is 5380 out of 10407 passing the prior-T5 trigger condition (cf. Table 5.1).

Applied cut	$\theta \leqslant 55^{\circ}$	No saturation	posterior-T5	Fitted β
No. Events	10290	9077	9077	5380

TABLE 5.1: Summary of the number of remaining events after every applied cut.

Optimal Distance

Although the `LDFFinder` configuration allows one to set D_{ref}, the module includes the possibility to calculate r_{opt} as soon as the LDF fit has been performed. To calculate the optimum distance, the β estimate is varied within its uncertainty a number of times, `<RoptN>`, which has been fixed to 10, and a new LFD fit is attempted. The point r_{opt} is defined as the distance from the shower axis corresponding to the crossing point of these resulting fitted LDFs.

Fig. 5.3 shows the distribution of r_{opt} computed in this way, for all the chosen reference distances. As β is not estimated from the parametrisation in Eq. 4.5, and hence is independent of $S_{r_{\text{opt}}}$, this distribution is independent from the choice of the reference distance. However, one can

observe a double-peak feature: the most probable value occurs at $\simeq$ 270 m together with a second peak at $\simeq$ 250 m.

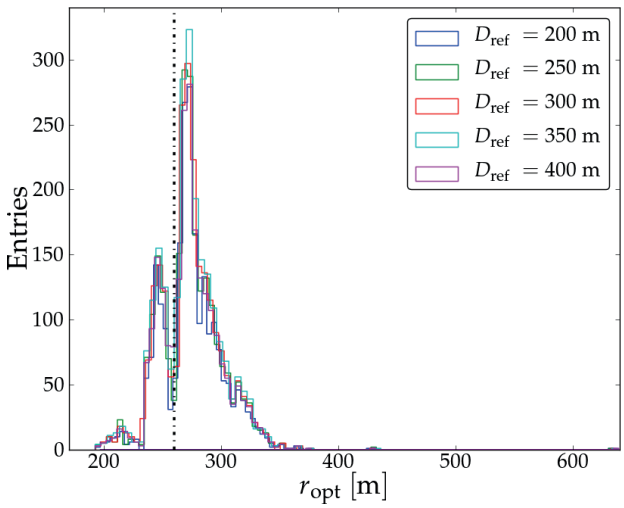

FIGURE 5.3: r_{opt} distribution of reconstructed events with fitted LDF-slope β for all chosen reference distances.

To investigate this feature, the dataset reconstructed with D_{ref} (= 200 m) has been first split into two samples around the local minimum between the two peaks at $\tilde{r}_{\mathrm{opt}} = 260$ m, indicated by the dash-dotted line in Fig. 5.3. After that, the shower-core positions for the two samples are shown in the two-dimensional histograms in Fig. 5.4, together with the AERAlet LS positions (black squares). The top (bottom) panel of this figure shows the core-position distribution for the sample with $r_{\mathrm{opt}}[j] > \tilde{r}_{\mathrm{opt}}$ ($r_{\mathrm{opt}}[j] \leqslant \tilde{r}_{\mathrm{opt}}$), for the j-th event. As one can see in the bottom panel, a high events-per-bin density is observed between the central tank 1764 and the LS 98, which has been installed for technical reasons closer to the central tank than the other LSs of the hexagon ($\approx$ 70-80 m less). This asymmetry in the array geometry is more accentuated if LS 98 is the station with the second highest signal. As shown in Ref. [88], r_{opt} depends on the mutual distance between the tanks, and it approaches this distance as much as the shower core falls close to a station.

Another way to look at the r_{opt} distribution is shown in Fig. 5.5. Here, the optimum distance is plotted versus the distance of the core to the highest-signal station $r_{S_{\mathrm{max}}}$. The black points represent the average r_{opt} in each $r_{S_{\mathrm{max}}}$-bin, the error bars correspond to 1σ spread in that bin, whilst the two-dimensional histogram in the background shows the event density. In the same figure, the two lines are respectively drawn at 250 m

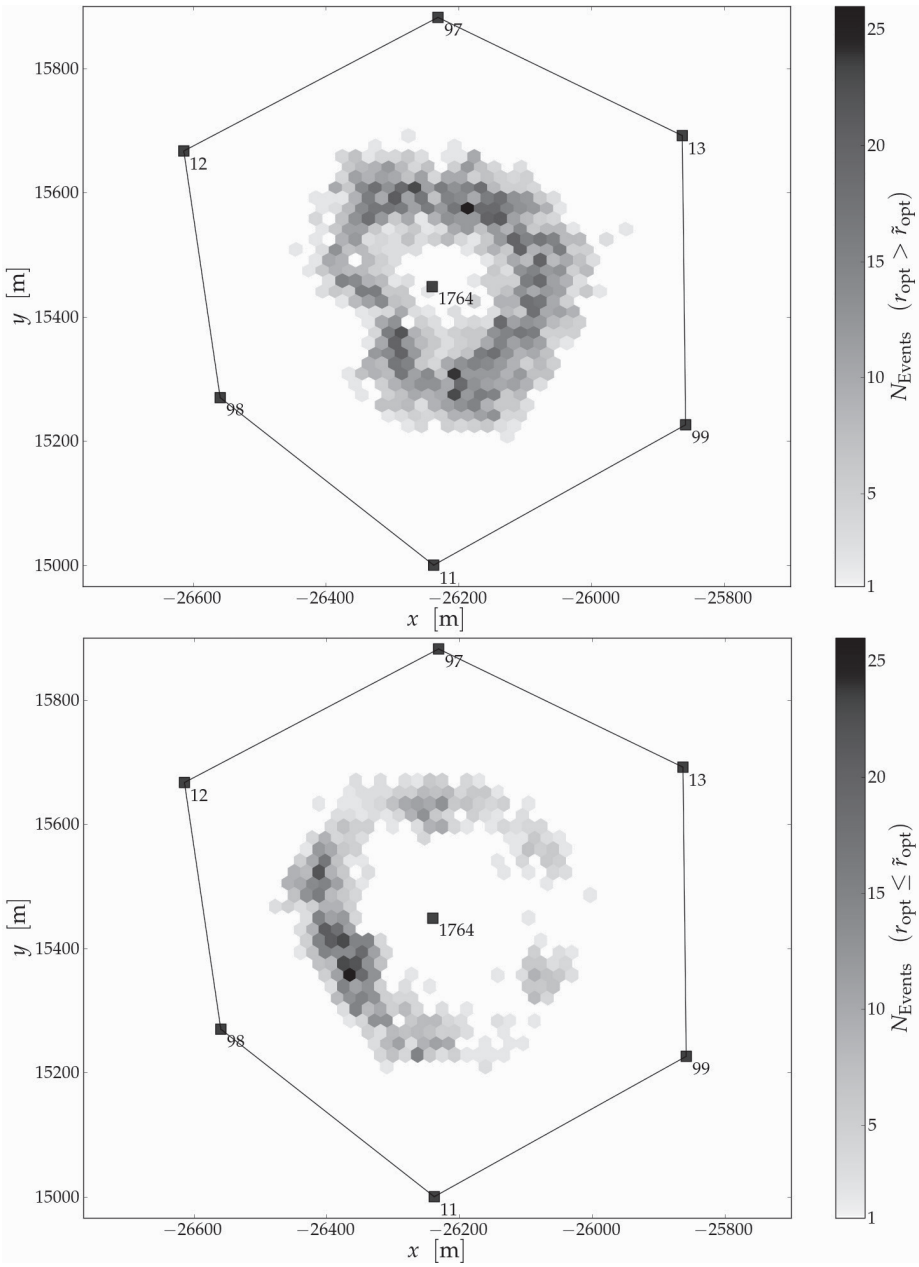

FIGURE 5.4: Shower-core positions of the reconstructed events with fitted LDF-slope β and $D_{\text{ref}} = 200$ m. The dataset of reconstructed events has been split into two samples around $\tilde{r}_{\text{opt}} \equiv 260$ m. *Top panel:* core-position distribution for events with $r^j_{\text{opt}} > \tilde{r}_{\text{opt}}$. *Bottom panel:* core-position distribution for events such that $r^j_{\text{opt}} \leqslant \tilde{r}_{\text{opt}}$.

(solid) and at 270 m (dashed). It is evident how the tail of the r_{opt} dis-

tribution rises for events where the highest-signal station is very close to the shower core, $r_{S_{\max}} \lesssim 100$ m. These events have a peculiar LDF, very steep close to the core, due to the station configuration: the highest-signal station very close, and all the other stations approximately at the same distance. At the same time, the bulk of the distribution concentrates around 270 m: looking at the binned values, the average distance where the slope fluctuations are minimal is 270 m. Therefore, in what follows, we shall consider $r_{\text{opt}} \equiv 270$ m and $S_{r_{\text{opt}}} \equiv S_{270}$. Table 5.2 summarizes the relation between array spacing and optimal distance for the Regular, Infill, and AERAlet arrays at the Observatory.

	Array Spacing [m]	r_{opt} [m]
Regular	1500	1000
Infill	750	450
AERAlet	433	270

TABLE 5.2: The optimal distance parameter r_{opt} for different array spacings.

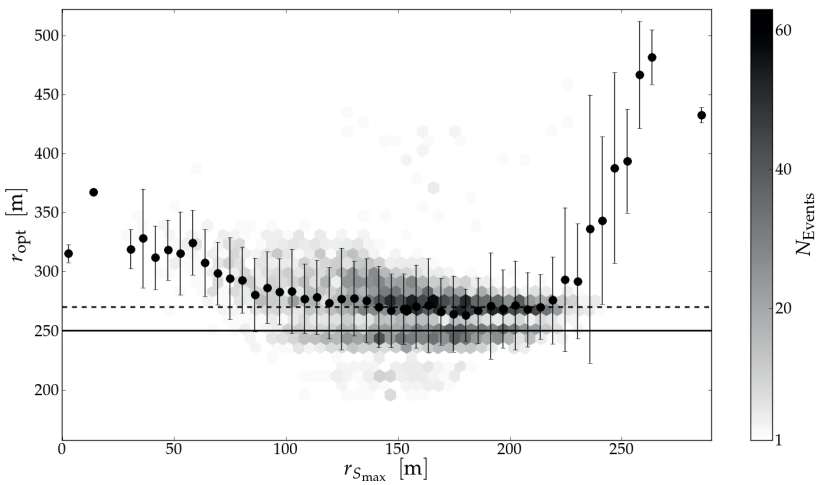

FIGURE 5.5: The optimal distance r_{opt} as a function of the distance of station with the maximum signal $r_{S_{\max}}$. The black circles represent the average value of r_{opt} values in that particular bin, the error bars their 1σ spread around the average, and the background histogram the events-per-bin density (bins with a single data point are not visible). Finally, the lines are respectively drawn at 250 (solid) and 270 m (dashed).

Slope Parametrisation

Once that $r_{\rm opt}$ has been fixed, the complete dataset has been reconstructed again as the optimal distance found was none of the $D_{\rm ref}$ considered. In what follows, this dataset has been used to determine the parametrisation in Eq. (4.5) adopting a *multi-event* fit approach rather than an *event-by-event* fit approach. In the latter case, one would consider to minimize the sum over the selected events, $\sum_i(\beta_i - \hat{\beta}_i)^2/\sigma^2[\beta_i]$, of the (squared) differences between the fitted LDF slope β_i and the model-provided value $\hat{\beta}_i = \hat{\beta}(\log S^i_{270}, \theta_i)$ given by Eq. (4.5), weighted by the associated variances $\sigma^2[\beta_i]$.

Alternatively, in a multi-event approach, one can fit the slope model for the global normalized LDF, i.e. the LDF obtained by collecting together the measured signals from all the selected events. For this purpose, each j-th measured signal of the i-th event, S^i_j, is divided by the LDF normalization factor $S^i_{r_{\rm opt}}$. Therefore, the main advantage of the multi-event-fit method is that the slope model in Eq. 4.5 can be fitted by exploiting a much larger number of data points than what it would be for the event-by-event approach. Indeed, from the 5380 selected events the amount of measured signals is 31833.

However, these signals need to pass an additional quality cut, known as *expected signal cut*. The aim of this cut is to remove a possible bias present in the data. Such a bias is due to upward fluctuations, around the trigger-threshold region, and downward fluctuations, around the tank-saturation region, of the measured signal with respect to the *true* signal. At the same time, it is not convenient to cut on the measured-signal values, because such a cut will introduce a further bias: close to the trigger threshold, the upward-fluctuating stations will be kept but the downward-fluctuating ones are rejected, regardless of what would have been the true signal (and vice versa close to the saturation region). Therefore, to avoid any bias we applied the expected signal cut: $5 \leqslant \hat{S}_j \leqslant 200$ VEM. The j-th expected signal $\hat{S}_j$ has been calculated by the signal model in Eq. (4.4), with r_j, S_{270}, θ retrieved from the single-event reconstruction, and the β-model parameters from a previous analysis on a smaller dataset.

After this cut, we are left with 23815 signals as input for the multi-event fit. For the fit, we have used a least-squares method to determine the set of parameters (a, b, c, d, e, f) of the β model as specified in Eq. (4.5), that

minimizes the objective function:

$$\chi^2 = \sum_{j=1}^{N_{\text{signals}}} \left\{ \frac{S_j - S_{270}^j \cdot f_{\text{LDF}}(r_j, \hat{\beta}(\theta_j, \log_{10} S_{270}^j | a, b, c, d, e, f))}{\sigma[S_j]} \right\}^2 . \quad (5.1)$$

For the minimization of the χ^2, and for all subsequent target functions, the KMPFIT module of the Kapteyn package [96] has been used. The minimum $\chi^2_{\text{red}} \simeq 1.28$ is reached for the set of parameters specified in Tab. 5.3.

Parameter	a	b	c	d	e	f
Value	-1.6	-0.5	-1.1	0.6	0.9	-0.4
Error	0.6	0.4	1.0	0.7	0.4	0.3

TABLE 5.3: Parameters of the LDF-slope (β) model, specified in Eq. (4.5), for the 433 m SD Infill array.

Before evaluating the goodness of this result, which will be tackled in the next Section, it is important to notice that, as mentioned in Section 4.2, if the slope is fixed by the model in Eq. (4.5), the uncertainty of this slope is fixed as well, and it is given in Offline by the slope-uncertainty model

$$\frac{\hat{\sigma}_\beta}{\beta} = -p_0 \exp\{-p_1 \log_{10} S_{270}\} \quad (5.2)$$

Evidently, the parameters of this model need to be optimized as well to reproduce the returned uncertainty σ_β, when β cannot be fitted. Though this optimization cannot be done in a multi-event approach, one can perform an event-by-event minimization for the model in Eq. (5.2):

$$\sum_{i=1}^{N_{\text{events}}} \frac{\left\{ \frac{\sigma_{\beta_i}}{\beta_i} - \frac{\hat{\sigma}_\beta}{\beta}(\log_{10} S_{270}^i | p_0, p_1) \right\}^2}{w[\sigma_{\beta_i}/\beta_i]}, \quad (5.3)$$

where β_i and σ_{β_i} are the fitted slope and its corresponding uncertainty as given by the minimization algorithm in Offline. Therefore, it is only needed to estimate the weight $w[\sigma_{\beta_i}/\beta_i]$ for the relative uncertainty of the slope parameter β. This weight has been calculated as the variance of the

ratio σ_β/β, by the error-propagation formula:

$$w[\sigma_{\beta_i}/\beta_i] \equiv \mathrm{Var}\left[\frac{\sigma_{\beta_i}}{\beta_i}\right] = \frac{1}{\beta_i^2}\sigma_{\beta_i}^2 + \left(-\frac{\sigma_{\beta_i}}{\beta_i^2}\right)^2 \mathrm{Var}[\sigma_{\beta_i}^2], \tag{5.4}$$

where $\mathrm{Var}[\sigma_{\beta_i}^2]$, the variance of the sample variance σ_β^2, needs to be estimated. For this purpose, for each i-th event, the distribution $\mathcal{N}(\beta_i, \sigma_{\beta_i})$ is generated with $M = 10^3$ samples, and being $\sigma_{\beta_i}^2$ the sample variance by definition, then one has [97]:

$$\mathrm{Var}[\sigma_{\beta_i}^2] = \frac{1}{M}\left(\mu_4 - \frac{M-3}{M-1}\sigma_{\beta_i}^4\right), \tag{5.5}$$

where μ_4 is the fourth sample moment for the generated distribution. In this way one can calculate the weights given by Eq. 5.4, to minimize the target function in Eq. (5.3). Finally, with a reduced χ^2 of 1.8, this process yields the parameter values in Table 5.4.

Parameter	p_0	p_1
Value	0.56	1.28
Error	0.01	0.01

TABLE 5.4: Parameters for the LDF-slope (β) relative-uncertainty model specified in Eq. (5.2).

Fig. 5.6 shows the two-dimensional distribution of the relative uncertainty of the slope $-\sigma_\beta/\beta$ versus $\log_{10} S_{270}/\mathrm{VEM}$, compared to the model for the parameter realization above (dashed curve). The same figure shows, a binned profile of the data together with the corresponding fit of the same model for the centroids of each bin, weighted by the error bar squared (solid curve). The two curves agree nicely with each other and with the data, except at low S_{270}. Here, the slope uncertainty, and correspondingly the variance of the relative uncertainty, is larger than this variance at high S_{270}, so these data points weight less in the objective function of Eq. (5.3), which is being pulled down in the un-binned case. However, as we shall see in the next Section, at low shower sizes (read energies) the detector is not efficient, and the recorded showers are biased by upward fluctuations.

5.2.4 Uncertainty and Bias Analysis

To evaluate the goodness of the result given by the least-squares multi-event fit above, one can look at the distribution of the residuals, i.e. the

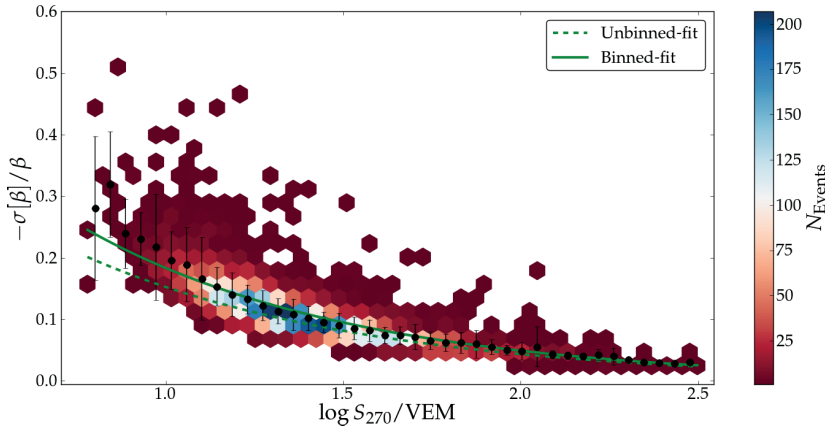

FIGURE 5.6: Comparison against the data of the slope-uncertainty model in Eq. (5.2) with the parameter realization specified in Tab. 5.4.

weighted difference between the measured signal, S_j and the model-predicted value $\hat{S}_j$:

$$\mathrm{Res}(S_j) = \frac{S_j - \hat{S}_j}{\sigma[S_j]}. \tag{5.6}$$

For an unbiased estimation of the parameters, one would expect normal distributed residuals centred at zero with unitary variance, as in this case the difference between the measurement and the model would be of the order of the measurement uncertainty. Fig. 5.7 shows the density of the residuals separated into five $\log_{10} S_{270}/\mathrm{VEM}$ bins. As can be noticed, the first moment of the PDF in each bin, indicated by the dashed line, drifts from low to high values from the lowest to the highest bin. This behaviour suggests the presence of a bias in our estimation process, more evident in the lowest and in the highest reference-signal bin.

As β is a function of $\log_{10} S_{270}/\mathrm{VEM}$, it seems that the parameters in Tab. 5.3 are less optimal for very-low and very-high energy events. On the one hand, as the AERAlet energy threshold has not been determined yet, the low-energy events are likely to have an energy below this threshold, and only upward-fluctuating signals are recorded. On the other hand, at the highest energies, due to the small spacing between the AERAlet stations, only the very central part of the shower front is sampled. As a consequence, the shower-size parameter S_{270} determined in Offline is likely to be estimated wrongly.

The previous statement can be easily verified by introducing a correction factor $\mathcal{K}$, in front of the normalization factor S_{270} of the LDF model

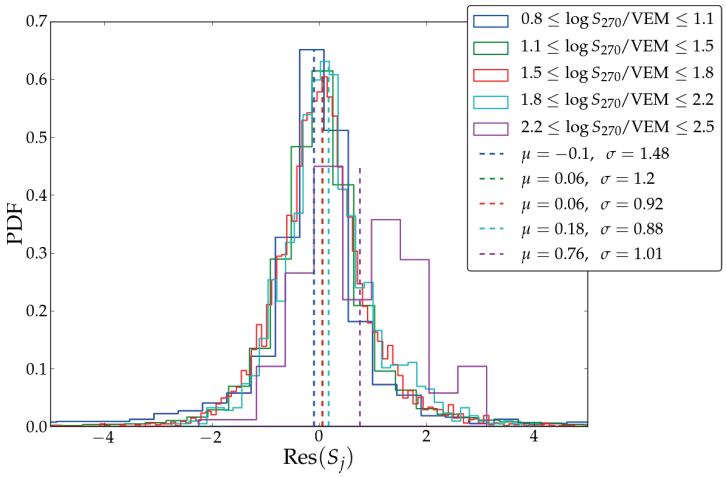

FIGURE 5.7: Distribution for the weighted difference between the measured signal, and the expected signal given by Eq. (4.4) with β model in Eq. (4.5) fixed by the parametrisation in Tab. 5.3. The data have been split into five reference-signal bins, and the dashed lines indicates the first moment of the fitted Gaussian p.d.f. (not shown for clarity) for each of these bins (see legend).

of Eq. (4.4). The modified LDF model reads:

$$\hat{S}'_l = \mathcal{K} \cdot S_{270} \left(\frac{r_l}{r_{\text{opt}}} \right)^{\bar{\beta}} \left(\frac{r_l + r_{\text{scale}}}{r_{\text{opt}} + r_{\text{scale}}} \right)^{\bar{\beta}}, \tag{5.7}$$

where $\bar{\beta}$ is the slope calculated from the model in Eq. (4.5), by means of the parameters in Tab. 5.3. Evidently, if the LDF model correctly reproduces the measured signals, the normalization factor S_{270} needs no correction, and the value of $\mathcal{K}$ will simply be equal to one. However, if the LDF fit is actually biased, the model cannot reproduce the measured signals equally well. In this case, the correction factor $\mathcal{K} \gtrsim 1$ if the signals (and so S_{270}) are underestimated by the model, and $\mathcal{K} \lesssim 1$ if the signals are overestimated.

Now, a way to determine $\mathcal{K}$, is to split the dataset in several $\log_{10} S_{270}$ bins, to evaluate the magnitude of this correction with respect to the reference-signal parameter S_{270}. This binning has been performed in two ways: regular (evenly-spaced) and equally filled-bins in $\log_{10} S_{270}/\text{VEM}$. For each i-th bin, a least-squares multi-event fit has been performed to

obtain $\mathcal{K}$, and its uncertainty $\sigma_{\mathcal{K}}$, minimizing

$$\sum_{l=1}^{N^i_{\text{signals}}} \left\{ \frac{S_l - \mathcal{K} \cdot \hat{S}_l}{\sigma[S_l]} \right\}^2, \qquad \text{for} \quad i = 1, \ldots, N_{\text{bins}}. \tag{5.8}$$

The uncertainty $\sigma_{\mathcal{K}}$ is the one returned by the minimization algorithm, however, in the target function above the uncertainty on S_{270} has not been taken into account.

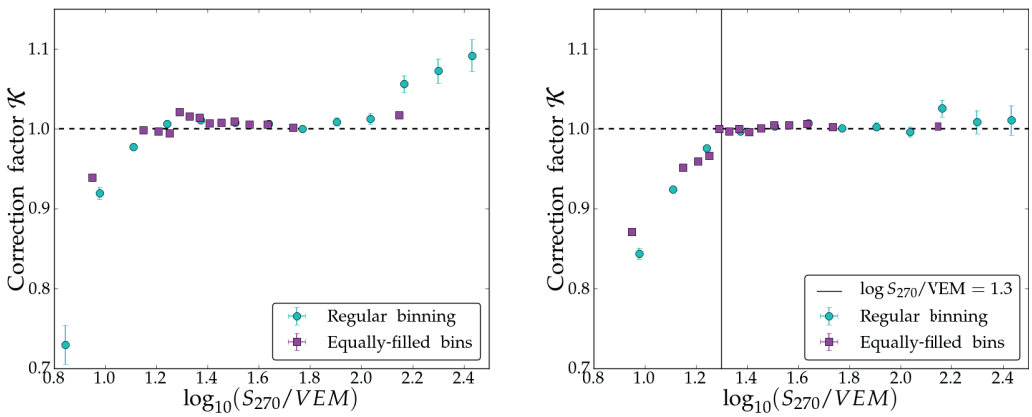

FIGURE 5.8: The fitted correction factor $\mathcal{K}$ per $\log_{10} S_{270}$/VEM bin. The fit is repeatesd for two types of binning the data: bins of equal widths (cyan dots) and bins containing an equal amount of signals (magenta squares).

Fig. 5.8-*left* shows the result of such a procedure; the fitted value $\mathcal{K}_i \pm \sigma_{\mathcal{K}_i}$ is plotted versus the centre of the i-th bin, for the two different binning procedures. As can be seen, the fit is indeed biased at very low and very high reference signals, i.e. the model with the parameters in Tab. 5.3 does not accurately describe the lateral fall-off of the signal. Evidently, in absence of an energy threshold, the efficiency of the detector is playing a role at the lowest energies, where the reconstruction is clearly biased by upward fluctuations, resulting in a systematic overestimation of S_{270}.

The step visible at $\log_{10} S_{270}/\text{VEM} \simeq 1.3$, in Fig. 5.8-*left*, suggests that to obtain an unbiased estimation of the β parameters, an additional (conservative) cut to our selected events can be applied requiring $\log_{10} S_{270}/\text{VEM} > 1.3$, which mimics an energy-threshold cut. This *anti-bias* cut reduces the number of events from 5380 to 3187, and the number of selected signals available for the fit to 15802. On this reduced dataset, the multi-event fit of the original model (i.e. with no correction factor) has been repeated, the minimum $\chi^2_{\text{red}} \simeq 1.0$ being reached for the set of parameters shown in

Table 5.5.

Parameter	a	b	c	d	e	f
Value	−1.9	−0.4	−0.7	0.6	0.7	−0.3
Error	1.2	0.8	2.0	1.3	0.8	0.5

TABLE 5.5: Parameters for the β model specified in Eq. (4.5), determined by the multi-event fit, after the additional anti-bias cut $\log_{10} S_{270}/\mathrm{VEM} > 1.3$ has been applied at the event level.

Then, one can first re-estimate $\mathcal{K}$, in every $\log_{10} S_{270}/\mathrm{VEM}$ bin, to check the performance of the model above the cut, using the parametrisation above for the minimization in Eq. (5.8). Fig. 5.8-*right* shows the result of this procedure. Above the cut, indicated in this Figure by the vertical black-solid line, $\mathcal{K} \simeq 1$, which means that the signal model with the parametrisation in Tab. 5.5 works better than the set of parameters in Tab. 5.3.

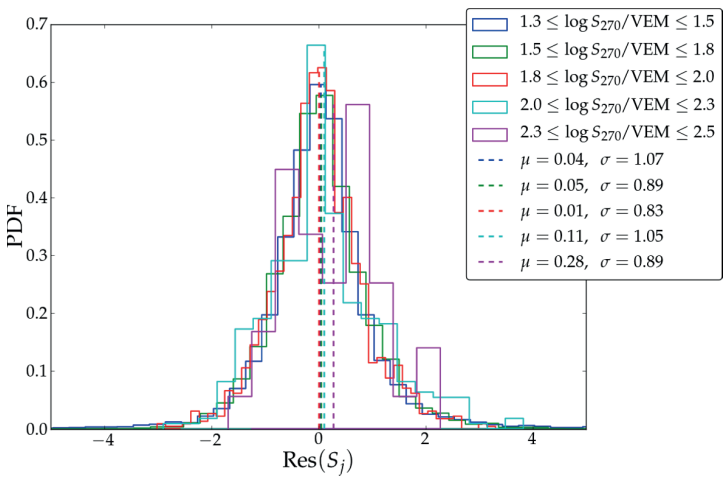

FIGURE 5.9: Same as Fig. 5.7 but for the parameters of Table 5.5.

Finally, one can look at the distribution of the residuals given by Eq. (5.6), to check if the multi-event estimation of the parameters is now unbiased. The residual distribution, into five $\log_{10} S_{270}/\mathrm{VEM}$ bins, is shown in Fig. 5.9 together with the first moment of the estimated density indicated by the dashed line for each bin. Comparing to the distributions in Fig. 5.7, and to their first moments, a significant improvement has been achieved, though in the last signal-reference bin the parameters are not yet optimal. This last bin only contains 81 events at the highest energies, for which we know that the shower front is sampled only close to the

core. The lateral fall-off of the signal is more steep close to the core than at far distances, so for high-energy showers the tail of the LDF cannot be correctly described.

Nevertheless, the new set of parameters can be safely used to reconstruct the LDF for the AERAlet array. Fig. 5.10 shows the average residuals as a function of the station distance from the shower core, for the parametrisation in Tab. 5.3 (*top panel*) and for the parametrisation in Tab. 5.5 after the anti-bias cut (*bottom panel*). As it can be seen, in the latter case, the average residual is basically zero with a unitary standard deviation, which means that the difference between the measured value and model-provided value is of the order of the signal uncertainty. The only exception is very close to the core, where the spread of the data points is wider than anywhere else. Although the spread is not as wide as for the previous case, here the distance of the station is more affected by the core-location uncertainty. As a consequence, the model-provided signal value will suffer from this worse accuracy in the distance determination, especially in this region, where a small change in distance corresponds to a large change in the computed signal.

Certainly, the large uncertainties of the new parametrisation do not go unnoticed. On the one hand, as the number of total data points is reduced by about 30% after the cut, the estimation is less precise. On the other hand, this result may be a hint that something else is going on.

In a previous study [98] the estimation of the β-model parameters has been investigated in detail. First, the parameters were estimated by performing a multi-event least-squares fit as in Eq. (5.1). Second, a MC simulation was launched to estimate the parameter p.d.f., by artificially fluctuating the measured signals, within their uncertainty, and each time repeating the multi-event fit. The estimated density was a normal p.d.f. for every parameter, but the first central moment was not corresponding to the parameters minimizing the χ^2 in the first place, and $\chi^2_{\mathrm{MC}} > \chi^2$. Third, as the parameters found in the first place consist of a minimum for the objective function in Eq. (5.1), one can expand this function in Taylor series around the minimum and consider the linear approximation of the LDF model. Then, for such a linear model, one can use the well known expression for the Minimum Variance Unbiased (MVU) estimator, which ultimately depends on the initial guess of the parameters. Finally, iterating this method, this resulted in a new set of parameters with $\chi^2_{\mathrm{MVU}}/\mathrm{ndof} \sim 0.8$. Essentially, as the parameters of each of these sets were correlated, it was concluded that several minima exist in the χ^2 landscape (a six-dimensional parameter space). However, the different parameter

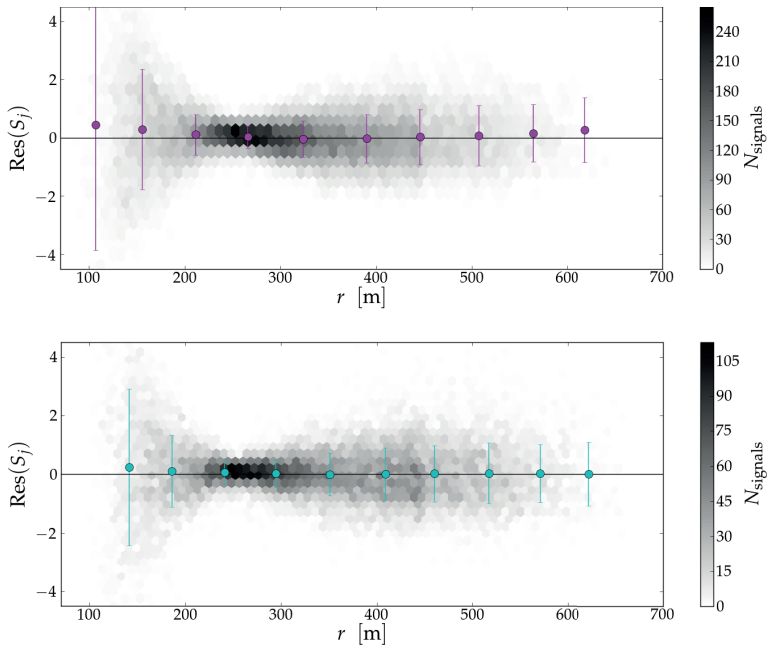

FIGURE 5.10: Average residuals, Eq. (5.6), as a function of the station distance from the core. Each point represents the mean of the residuals in that bin, and the error bar its standard deviation. The points have been over-imposed to the actual two-dimensional distribution showing the spread of the signals as a function of distance. *Top panel:* β fixed by the parametrisation in Tab. 5.3. *Bottom panel:* after applying the $\log_{10} S_{270}/\text{VEM} > 1.3$ cut and fixing β by the parametrisation in Tab. 5.5.

realizations have been used to estimate the LDF systematic uncertainty $\sigma^{\text{sys}}_{\text{LDF}}$, associated with the choice of one particular parametrisation.

As the signal model depends mainly on the shower-plane distance, at each distance r_i one can evaluate the magnitude of the difference between the model-provided values for two different parameter sets

$$\Delta S(r_i) = |S_{\text{MVU}}(r_i) - S_{\text{MLS}}(r_i)| \qquad i = 1, \ldots, N_{\text{signals}}, \tag{5.9}$$

where $S_{\text{MVU}}(r_i)$, $S_{\text{MLS}}(r_i)$ indicate respectively the model-provided signal with the MVU and the multi-event least-squares fit parameters (MLS). Finally, the systematic uncertainty has been estimated as the most probable value of this empirical distribution, $\sigma^{\text{sys}}_{\text{LDF}} \simeq 0.14$ VEM.

5.3 Energy Calibration

As mentioned in Section 4.2.1, the signal at the optimal distance S_{270} depends on the energy E and on the zenith angle θ of the primary CR. Such a zenith-angle dependency is mainly due to the atmospheric attenuation: as the amount of traversed atmosphere by the shower increases (with θ) the electromagnetic component is largely absorbed, and the measured particle density (and accordingly the measured signals) decreases. This effect is clearly independent from the shower azimuth angle, as the amount of atmosphere does not change with it.

In the next Section, we shall make use of the Constant Intensity Cut method to obtain the zenith-independent energy estimator $S_{\theta_{\rm ref}}(E)$. Only then, it will be possible to calibrate this energy estimator with a reference energy measurement.

For the present calibration we used CR events that have been detected both by the 750 m Infill array and by AERAlet. Though the preferred calibration method would have been to use the calorimetric energy measurement by FD, we faced two major problems that prevented this possibility. On the one hand, we found very few events in coincidence between AERAlet and the HEAT/Coihueco eyes, partially due to the limited duty cycle of FD. On the other hand, those coincident events were extremely unlikely to pass the strict quality cuts for a reliable energy measurement.

Therefore, a set of events simultaneously observed by AERAlet and the Infill arrays, has been reconstructed using the two arrays independently, and before performing the actual calibration, the two geometrical reconstructions are compared in Section 5.3.2.

5.3.1 Energy Estimator for AERAlet

The basic assumption is that above the trigger-threshold energy, $E_{\rm thr}$, the cosmic-ray flux J is isotropic. Therefore the energy and zenith-angle dependencies of $S_{270}(E,\theta)$ can be factorized.

The flux of cosmic rays J as a function of energy is given by [99]:

$$J(E) = \frac{d^4N}{dE\,dA_{\rm eff}\,d\Omega\,dt}, \tag{5.10}$$

where N is the number of cosmic rays with energy E incident on an effective-surface element $dA_{\rm eff} = \cos\theta\, dA$, within a solid angle $d\Omega = d\cos\theta\, d\phi$

and time dt. Hence, the condition of an isotropic flux reads:

$$\frac{d}{d\theta}J(E) = 0 \quad \Rightarrow \quad \frac{d}{d\theta}\left(\frac{d^4N}{dE\,d\cos^2\theta\,dA\,dt}\right) = 0, \tag{5.11}$$

where $2\cos\theta\, d\cos\theta = d\cos^2\theta$ has been used. The order of the derivatives can be exchanged for zenith-independent variables, and the previous condition is then equivalent to

$$\frac{d}{d\theta}\left(\frac{d^2N}{dE\,d\cos^2\theta}\right) = 0 \quad \Rightarrow \quad \frac{d^2N}{dE\,d\cos^2\theta} = \text{constant}. \tag{5.12}$$

The equation above can be integrated above the trigger-threshold energy to define the intensity I of events

$$\int_{E_{\rm thr}}^{\infty} dE\,\frac{d^2N}{dE\,d\cos^2\theta} =: \left.\frac{dI}{d\cos^2\theta}\right|_{E=E_{\rm thr}} \cong \frac{\Delta I}{\Delta(\cos^2\theta)} = \text{constant}, \tag{5.13}$$

i.e. for $E > E_{\rm thr}$, one expects the same amount of events in equally-spaced $\cos^2\theta$ intervals, or in other words the event *intensity* (number of events divided by the interval width) is a constant. This condition can be used to infer the correction needed to disentangle energy and zenith-angle dependencies of the signal at the optimal distance S_{270}. Indeed, splitting the S_{270} distribution in several equally-spaced $\cos^2\theta$ bins, as shown in Fig. 5.11, one observes that the same number of events is detected at high S_{270} as the zenith-angle increases, which confirms that the zenith-angle dependency of S_{270} can be factorized (cf. Eq (4.6)):

$$S_{270}(E,\theta) = S_{270}(E,\theta=\theta_{\rm ref})\cdot f_{\rm CIC}(\tilde{\theta}) = S_{\theta_{\rm ref}}(E)\cdot f_{\rm CIC}(\tilde{\theta}), \tag{5.14}$$

where the definition $S_{\theta_{\rm ref}}(E) := S_{270}(E,\theta=\theta_{\rm ref})$ is used. The function $f_{\rm CIC}(\tilde{\theta})$ is the correction function to be determined, with functional form:

$$f_{\rm CIC}(\tilde{\theta}) = 1 + a\tilde{\theta} + b\tilde{\theta}^2 + c\tilde{\theta}^3, \qquad \text{with} \qquad \tilde{\theta} = \cos^2\theta - \cos^2\theta_{\rm ref}. \tag{5.15}$$

The reference angle $\theta_{\rm ref}$ is commonly taken in the centre of the $\cos^2\theta$ range considered, whose upper limit is fixed by the zenith-angle cut. Contrarily to the zenith-angle cut of the previous analysis, it was fixed to $\theta_{\rm cut} = 50^\circ$. The reason for decreasing this cut originated from a previous CIC analysis [100] with $\theta_{\rm cut} = 55^\circ$ and hence $\theta_{\rm ref} = 35^\circ$. The reference signal S_{35}, obtained after the derived correction function, was still showing a θ dependency for inclined low-energy showers. Evidently at low energies, the electromagnetic component is completely absorbed between

50° and 55°, and the correction function is not able to entirely take into account these strong attenuation effects. As a consequence, the zenith angle range of vertical events for AERAlet has been set to $[0°, 50°]$, which gives $\theta_{\rm ref} = 33°$.

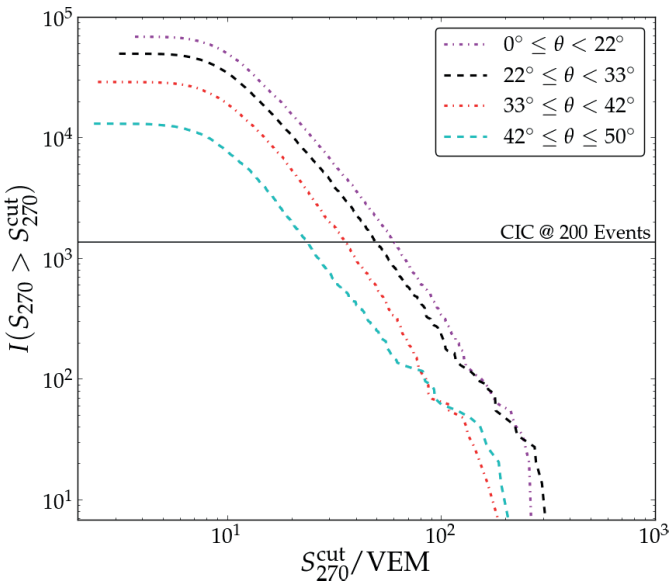

FIGURE 5.11: Intensity of events with $S_{270} > S_{270}^{\rm cut}$ in four angular bins of equal width in $\cos^2\theta$. For every $S_{270}^{\rm cut}$ on the x-axis, the number of events divided by the bin width is reported on the y-axis. The solid-horizontal line indicates the *constant intensity* at 200 events, which is achieved at small $S_{270}^{\rm cut}$ values as the zenith angle increases. Because of the atmospheric attenuation, at the ground level, inclined showers will contain less particles than vertical ones.

For the present analysis, the same dataset used before has been reconstructed again fixing β from the parametrisation for all events. During the reconstruction, events falling in the bad-SD periods have been rejected. These periods are known time ranges during which technical issues or weather phenomena affect the air-shower SD measurements. To not discard many events, as the condition for the standard posterior T5 can only be satisfied if the station with the highest signal is the central one, we decided to set the number of stations required for the posterior T5 to 3, while making sure that the central tank had a signal as well. Though this criterion could include not-entirely sampled showers, the core is still contained in a triangle of active stations, and the final reconstructed S_{270} still reliable, as confirmed by its relative-uncertainty (for the 95% of events $\sigma[S_{270}]/S_{270} < 0.19$, and for the 99% of events $\sigma[S_{270}]/S_{270} < 0.26$).

Fig. 5.11 shows the intensity of events with $S_{270} > S_{270}^{\rm cut}$ in four angular

bins of equal width in $\cos^2\theta$. The horizontal solid line indicates the constant intensity at 200 events, which is clearly not achieved for the same value of S_{270}^{cut} in the different angular bins. I. e. this cut value for the first bin corresponds to $S_{270}^{\text{cut}} \approx 80$ VEM, decreasing to 20 VEM for the last bin. For a zenith-independent energy estimator S_{33}, an equal intensity needs to be observed at all zenith angles.

This conclusion is equivalent to what is stated in Eq. (5.14) as the constant in Eq. (5.13) is fixed: for a certain intensity (cut) value, the function in Eq. (5.15) determines the correspondence between the different zenith-dependent cut values S_{270}^{cut} (cf. Fig. 5.11), and the single value of the signal at the reference angle. Indicating with j the zenith-angle-bin number, with θ_j its central value, one has symbolically

$$S_{270}^{\text{cut}}(\theta_j) = \mathfrak{s}_{33}^{\text{cut}} \cdot f_{\text{CIC}}(\tilde{\theta}_j|a,b,c) \quad \forall j \in \{1,2,\ldots,n_{\text{bins}}\}. \tag{5.16}$$

Here, the symbol $\mathfrak{s}_{33}^{\text{cut}}$ is introduced because for every value of the constant-intensity cut one can define a value of the signal at the reference angle by Eq. (5.16), to be distinguished from the final converted value S_{33}. Therefore, the parameters $\mathfrak{s}_{33}^{\text{cut}}, a, b, c$ can be obtained by minimizing the following objective function

$$\sum_{j=1}^{n_{\text{bins}}} \frac{\{S_{270}^{\text{cut}}(\theta_j) - \mathfrak{s}_{33}^{\text{cut}} \cdot f_{\text{CIC}}(\tilde{\theta}_j|a,b,c)\}^2}{\sigma^2[S_{270}^{\text{cut}}]}. \tag{5.17}$$

the $S_{270}^{\text{cut}}(\theta_j)$ values have been extracted slicing the data in $n_{\text{bins}} = 10$ equally-spaced $\cos^2\theta$ bins, and consequently applying a Constant-Intensity Cut at 200 events. However, the uncertainties $\sigma[S_{270}^{\text{cut}}]$ of these values are unknown. To adequately weight the minimization, one can estimate these uncertainties using a bootstrap method [101].

In short, for every zenith-angle bin, there is an empirical distribution of $S_{270}(\theta_j)$, from which the cut value is extracted. This distribution is completely empirical, i.e. the underlying p.d.f. describing how the $S_{270}(\theta_j)$ values are distributed in each bin is not known. As the observations are independent, this distribution can be approximated by "resampling with replacement" a number of times equal to the size of the original distribution. Then, to this approximate distribution, the same constant-intensity cut is applied to extract a $\mathcal{S}_{270}^{\text{cut}}(\theta_j)$ value. Repeating this operation $M = 10^3$ times, M cut values are extracted to construct the distribution $\{\mathcal{S}_{270}^{\text{cut},1}(\theta_j), \mathcal{S}_{270}^{\text{cut},2}(\theta_j), \ldots, \mathcal{S}_{270}^{\text{cut},M}(\theta_j)\}$ from which the variance can be estimated.

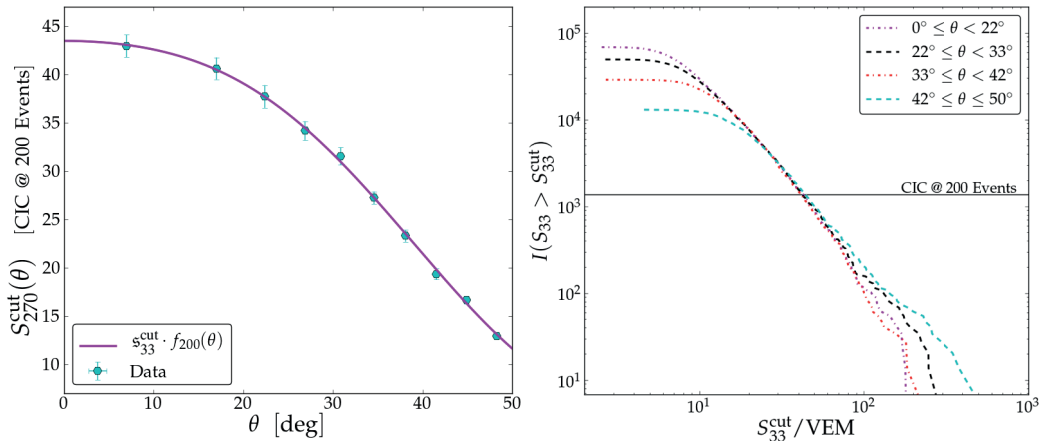

FIGURE 5.12: *Left panel:* Comparison between the S_{270} cut values obtained for a CIC at 200 events, and the corresponding fitted correction function in Eq. (5.18). *Right panel:* Intensity of events with $S_{33} > S_{33}^{\rm cut}$ in four angular bins of equal width in $\cos^2\theta$. For the i-th event, the zenith-independent reference signal S_{33}^{i} is obtained as $S_{270}^{i}/f_{200}(\tilde{\theta}^{i})$ (see text for details). For every $S_{33}^{\rm cut}$ on the x-axis the number of events divided by the bin width is reported. The solid-horizontal line indicates the *constant intensity* at 200 events.

The results of minimizing Eq. (5.17) are shown in Fig. 5.12, for the chosen CIC at 200 Events. In Fig. 5.12-*left*, the extracted cut values $S_{270}^{\rm cut}(\theta_j)$ are plotted versus the centroids of the corresponding $\cos^2\theta$ bins. Therefore, given the correspondence in Eq. (5.16), after the minimization the correction function

$$s_{33}^{\rm cut} \cdot f_{\rm CIC}(\tilde{\theta}_j | a, b, c) \equiv 28.9 \cdot f_{200}(\tilde{\theta} | a = 2.2, b = -0.6, c = -3.5), \qquad (5.18)$$

reproduces these values as shown by the solid-magenta line in this figure.

Thereafter, assuming that the correction function above is valid at any intensity, every measured S_{270} can be converted to the corresponding S_{33} value according to Eq. (5.14). Fig. 5.12-*right* shows the intensity spectra, in four $\cos^2\theta$ bins, for the attenuation-corrected reference signal $S_{33}(E)$, i.e. the number of events with S_{33} greater than the corresponding $S_{33}^{\rm cut}$ on the x-axis. As one can see, comparing with Fig. 5.11, the intensity corresponding to 200 events is now attained at the same reference signal $S_{33}(E)$, in each zenith-angle bin. In other words, except for the region below the trigger-threshold ($S_{33}^{\rm cut} \lesssim 25$ [VEM]), the θ and energy dependencies of the reference signal have been completely disentangled.

On the one hand, the choice of the CIC seems something arbitrary, on the other hand, it is evident from the intensity spectra shown in Fig. 5.12-*right*

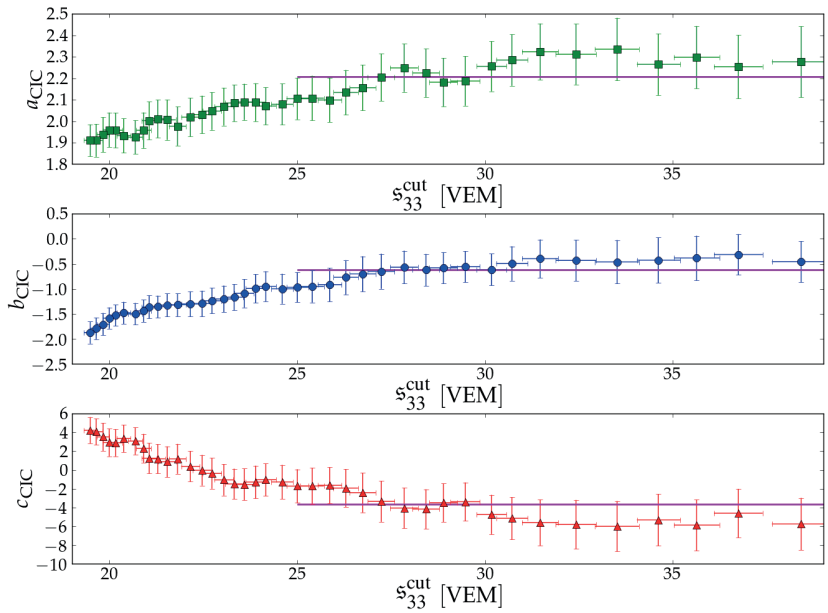

FIGURE 5.13: Variation of the fitted parameters a, b, c for the correction function in Eq. (5.15), as one varies the number of events at which the *CIC* is performed.

and from the definition in Eq. (5.13), that the detector is not efficient at the very low energies mainly because of the atmospheric attenuation. Therefore, the CIC method cannot be applied at high intensities for the reason above, and at low intensities due to the scarce event statistics. To avoid to arbitrarily picking up a number of events for the CIC, we have scanned the range of intensities corresponding to $[500, 100]$ events (in steps of 10), excluding these two regions.

Fig. 5.13 shows the estimated parameters by minimizing Eq. (5.17) for every CIC, indicated as $a_{\rm CIC}, b_{\rm CIC}, c_{\rm CIC}$, against the corresponding $s_{33}^{\rm cut}$ fitted value. In this figure, for each parameter, the solid-magenta line represents the fitted constant in the stable region, to finally obtain the values of a, b, c shown in Tab. 5.6. Lastly, these values will be used for the conversion of $S_{270}(E, \theta)$ in the AERAlet energy estimator $S_{33}(E)$, combining Eq. (5.14) with Eq. (5.15):

$$S_{33}^{i} = \frac{S_{270}^{i}}{f_{\rm CIC}(\tilde{\theta}_i | a, b, c)} \qquad i = 1, 2, \ldots, N_{\rm events} \tag{5.19}$$

As below the trigger threshold the number of detected events depends strongly on the zenith angle, the efficient region for AERAlet can be identified by checking for which S_{33} the intensity in different zenith-angle

Parameter	a	b	c
Value	2.21	−0.63	−3.7
Error	0.03	0.08	0.5

TABLE 5.6: Final parameters for the CIC correction specified in Eq. (5.14), see text for details about their determination.

bins starts to be consistent with a constant. To check this assumption, one can set up an iterative process starting from a very small value of S_{33}^{cut} [87].

Similarly to what was done for the intensity spectra, all events above this cut value are taken into account, and consequently separated into N_{bins} zenith-angle bins. If the constant-intensity assumption holds, the number of events per bin n_i is uniformly distributed. Therefore, a constant can be fitted to the entries n_i with weights $\sqrt{n_i}$ ($i = 1, 2, \ldots, N_{\text{bins}}$), and corresponding χ^2_{red}. Then, the S_{33}^{cut} is increased by 0.5 VEM, and the procedure above repeated, up to a not-too-large value of S_{33}^{cut} where the event statistics starts to be poor.

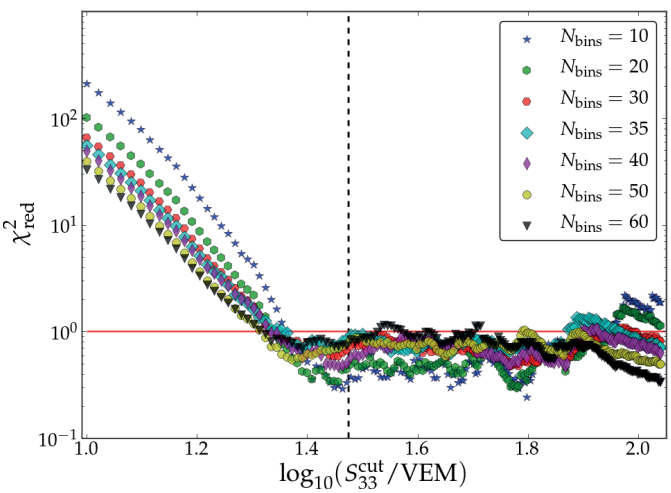

FIGURE 5.14: Consistency check for the constant-intensity assumption for events with $S_{33} > S_{33}^{\text{cut}}$. After separating the events in N_{bins} angular bins (see legend), every point in the plot is the resulting χ^2_{red} of a constant fit to the entries $\{n_i\}_{i=1}^{N_{\text{bins}}}$ for every S_{33}^{cut} on the x-axis.

Fig. 5.14 shows the results of the process just described for $N_{\text{bins}} \in \{10, 20, 30, 35, 40, 50, 60\}$. Indeed, the χ^2_{red} values need to be taken carefully, as it can be noted that N_{bins} can quantitatively influence the χ^2: few bins with a lot of entries will have large weights with the effect of lowering the χ^2

value. In this Figure, the red-solid line is drawn at $\chi^2_{\rm red} = 1$, which is reached for $\log_{10}(S^{\rm cut}_{33}/{\rm VEM}) \simeq 1.47$ indicated by the black-dashed line. It can be seen that the $\chi^2_{\rm red}$ drops below one and then rises again, which is a hint that other effects are affecting the efficiency.

Usually, the efficiency of the detector is accurately investigated by making use of simulations, as it depends on the mass composition as well in addition to the zenith-angle and energy dependencies. In the absence of complete simulations at these energies, an approximate trigger-threshold can be extracted from this plot, in correspondence to the black-dashed line, from which we have $S^{\rm thr}_{33} \simeq 30$ VEM.

Energy-Estimator Uncertainty

Once the conversion of the optimal-distance signal S_{270} to the estimator S_{33} has been established, one needs to calculate the uncertainty $\sigma[S_{33}]$. Basically, this has been estimated by propagating the uncertainty of S_{270} returned by the $\overline{\rm Off}\underline{\rm line}$ reconstruction (major contribution) using Eq. (5.14), to which we add the systematic contribution due to the LDF parametrisation, and the systematic contribution due the CIC method:

$$\sigma^2[S_{33}] = \left(\frac{\partial S_{33}}{\partial S_{270}}\right)^2 \sigma^2[S_{270}] + \overbrace{\sigma^2_{\rm LDF} + \sigma^2_{\rm CIC}}^{\sigma^2_{\rm sys}} \,. \tag{5.20}$$

The $\sigma_{\rm LDF}$ has been estimated at $\simeq 0.38$ by applying the same analysis in [98] (see discussion at the end of Section 5.2.4), to our dataset. The $\sigma^2_{\rm CIC}$ contribution has been taken, on an event basis, as the variance of the distribution constructed by calculating the possible values of S_{33} using the different sets of $f_{\rm CIC}$ parameters obtained varying the Constant-Intensity Cut as mentioned above. The different correction functions $f_{\rm CIC}(\tilde{\theta}|\, a_{\rm CIC}, b_{\rm CIC}, c_{\rm CIC})$ for the considered sets of parameters are shown in Fig. 5.15, as $S^{\rm cut}_{270}/S^{\rm cut}_{33}$ ratio. This contribution to the uncertainty depends on the zenith angle, as the correction vanishes at the reference angle $S^{\rm cut}_{270}/S^{\rm cut}_{33} \simeq 1$, whilst it is more accentuated at small and large angles, and so its variation.

As expected, the uncertainties $\sigma[S_{33}]$ are strongly energy correlated (being so the uncertainties $\sigma[S_{270}]$).

In Section 5.3.3, the following model has been fitted to the data, to realistically provide the uncertainty of S_{33} in the context of a toy Monte Carlo

(cf. Fig. 5.19-*right*)

$$\frac{\hat{\sigma}(S_{33})}{S_{33}} = c_0 + (c_1 \sec\theta + c_2)\frac{1}{\sqrt{S_{33}}} + (c_3 \sec\theta + c_4)\frac{1}{S_{33}}. \tag{5.21}$$

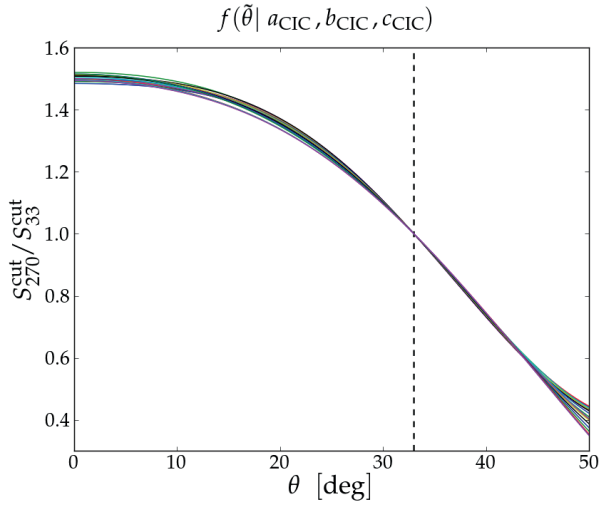

FIGURE 5.15: Variation of the correction function $f_{\rm CIC}(\tilde{\theta}|a_{\rm CIC}, b_{\rm CIC}, c_{\rm CIC}) = S_{270}^{\rm cut}/S_{33}^{\rm cut}$ using the different sets of fitted parameters obtained by varying the number of events at which the CIC is performed.

5.3.2 Event Selection and Reconstruction

To identify the set of CR events, simultaneously detected by the Infill and AERAlet arrays, a pre-selection step has been performed with the CDAS searching in the raw-data stream events occurring at equal GPS seconds but flagged with two different trigger algorithms (one of which was of course *AERA*). During this pre-selection, we realized that the central tank Kathy Turner (Id 1764) was missing from the Infill-detected events. For the current analysis, these events have been merged with the AERAlet events, adding to every Infill event the missing-station information from the corresponding AERAlet event.

After this merging, the raw data have been reconstructed twice with Offline, the first time selecting the Infill grid, and the second time selecting the AERAlet grid, through the `<GridType>` tag in the `SdEventSelector` configuration file. The AERAlet grid has been further enlarged by adding the first crown of Infill stations around the central tank Kathy Turner, namely SD stations 688, 1769, 1765, 734, 1773, 1622. As a matter

of fact, these stations have an approximate distance of 433 m from their AERAlet first neighbours, though they are not requested by the *AERA* trigger (see Section 3.4.1). Such an enlarged grid allows for a more precise reconstruction in Offline, and it has been possible to use them only because of this merging.

After the reconstruction, the following selection chain is needed to identify those CR events that actually have been recorded by both arrays. In fact, one cannot exclude that at the same GPS second another shower triggered the Infill array far from the AERAlet array.

1. The following quality cuts are applied to the Infill-reconstructed data:
   - Zenith-angle cut: $\theta_{\text{Infill}} \leqslant 55^{\circ}$.
   - Distance between Infill-reconstructed shower-core location and Kathy Turner < 300 m.
   - No saturated events.
2. The following quality cuts are applied to the AERAlet-reconstructed data:
   - Zenith-angle cut: $\theta \leqslant 50^{\circ}$.
   - No saturated events.
   - only posterior-T5 events.
3. Difference between the two reconstructed *core times* $< 1\ \mu$s.

After this selection, a total of 1452 events are left for the following analyses.

Geometrical-Reconstruction comparison

From this set of matched events, it is possible to estimate the angular and the core-position resolution for the AERAlet array. For this purpose, we have calculated for every event the opening angle η between the Infill-reconstructed and the AERAlet-reconstructed shower axis, and the magnitude of the ground-vector difference between the two reconstructed shower-core positions.

Fig. 5.16-*left* shows the distribution of such opening angles, and the corresponding fitted chi-squared PDF. Among all events, we have found seventeen events with $\eta > 10^{\circ}$, not shown in the plot. According to the Infill energy E_{Infill}, these events are very low-energy ones, for which the angular uncertainty is quite high. Indeed, up to this moment, no energy

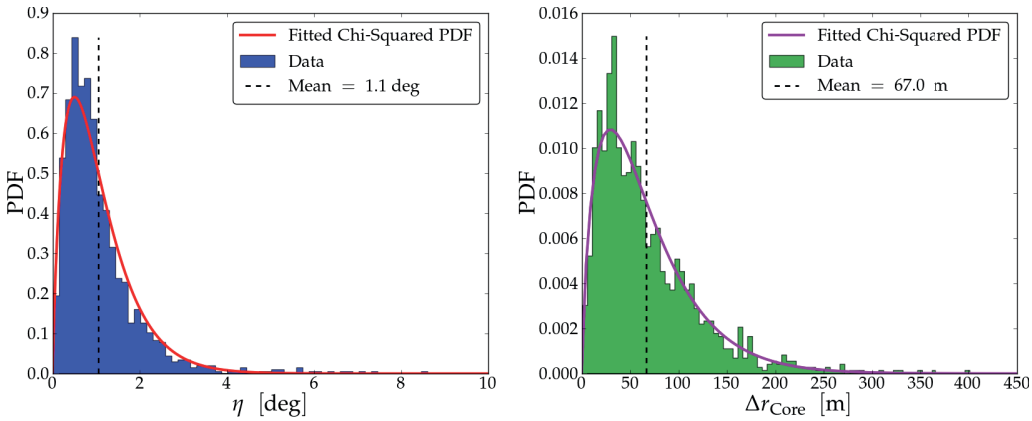

FIGURE 5.16: *Left panel:* Opening-angle η distribution between the Infill-reconstructed and the AERAlet-reconstructed shower axis, and the superimposed fit (red-solid line). The mean of the fitted PDF (black-dashed line) represents the estimated angular resolution for AERAlet. *Right panel:* Core-location difference Δr_{Core} distribution and corresponding fitted PDF (magenta-solid line). The mean of the PDF (black-dashed line) is the core-resolution estimate for AERAlet.

cuts have been considered to avoid that too many events are discarded. This also explains the long tails of the distribution. Nevertheless, the angular resolution of AERAlet can be safely estimated to be 1.1°, according to the mean of the fitted PDF, which nicely coincides with the 68% limit of the underlying empirical distribution. By the way, this value represents an upper limit to the angular resolution, as it is known that this quantity considerably improves with increasing θ, and with the number of stations participating to the reconstruction.

Similarly, Fig. 5.16-*right* shows the distribution of the core-location difference, and the corresponding fitted chi-squared PDF. The core-location resolution has been taken to be 67 m, according to the mean of the fitted PDF, which is very close to the 68% limit of the empirical distribution at 74 m.

In what follows, we shall make use of a sub-selection of these events, to calibrate the AERAlet energy estimator S_{33}, using couples $(E_{\text{Infill}}, S_{33})$. As mentioned above, E_{Infill} is the reconstructed energy by the Infill reconstruction, which is known to be fully efficient above 0.3 EeV [102] (although a less strict threshold at 0.2 EeV can be used safely [87]). Therefore, such an energy cut needs to be applied before attempting the calibration procedure. In this respect, although a likelihood method would be preferable, we shall make use of a least-squares method combined with an anti-bias cut, based on a MC study explained in the next Section.

The reason for this choice is mainly the lack of a suitable PDF to describe the trigger efficiency for AERAlet, and the existing intrinsic correlation between E_{Infill} and S_{33} due to the fact that the two arrays have at least one station in common (the central tank Kathy Turner). This correlation would require the addition of an extra unknown term in the construction of the Likelihood function.

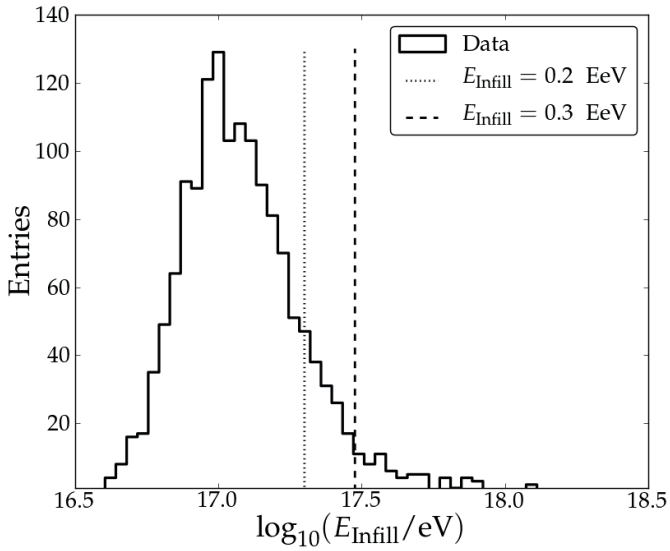

FIGURE 5.17: Energy distribution for the selected coincident events between AERAlet and the 750 m SD Infill array. The dashed and dotted lines indicate respectively the infill-energy threshold at 0.3 EeV and its less strict value at 0.2 EeV. The bins have been taken of equal size (0.04) in $\log_{10} E_{\text{Infill}}$.

5.3.3 Toy Monte Carlo

In the previous Section, we made a high-quality selection of CR events simultaneously measured by the Infill and AERAlet arrays. The final aim is to use these events to estimate the calibration coefficients necessary to convert the reference signal S_{33} into the primary energy of the CR initiating the air shower.

The energy distribution for this selection is shown in Fig. 5.17, where E_{Infill} is the energy measurement that will serve as reference. This Figure also shows, the official threshold at 0.3 EeV (magenta-dashed line) and the less strict threshold $\tilde{E}_{\text{Infill}} \equiv 0.2$ EeV (cyan-dashed line), which has been adopted for the rest of the analysis to gain more data points for the final calibration fit. Such a low threshold can be safely adopted as the

drop in the efficiency starts to be important for inclined showers, which are not considered because of the AERAlet zenith-angle cut at 50°.

Although the distribution of measured events peaks at $\approx 10^{17}$ eV, in the region below $\tilde{E}_{\text{Infill}}$, the Infill array is not fully efficient, upward-fluctuating signals have a high chance to be detected. Therefore, it is necessary to reject these events before performing any calibration fit. This rejection can introduce a bias if the relative uncertainties of S_{33} and E_{Infill} are not suitably taken into account [84]. Such a bias, for a two-dimensional linear fit, is minimum as the number of data points above and below the fit line is on average the same. If the relative uncertainties of S_{33} and E_{Infill} are *equal* and do not change as the energy changes, this minimum bias condition is achieved rejecting data points below a cut line *orthogonal* to the fit line.

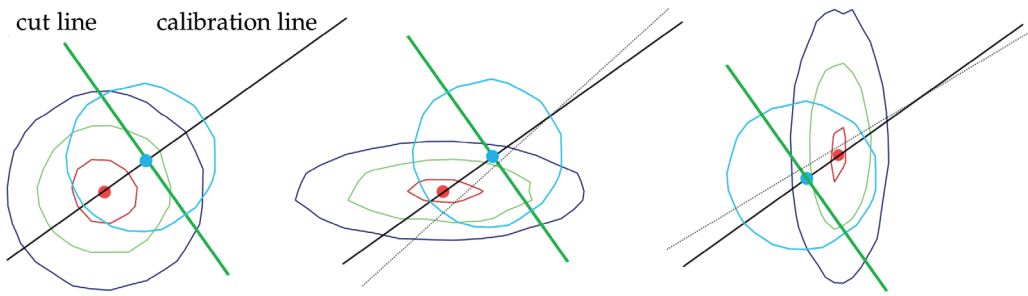

FIGURE 5.18: Schematic illustration of the possible bias introduced by an orthogonal cut line for the three different cases explained in the text. The cyan dot represents a point on the calibration line for which the 1σ contour (cyan line) appears as a circle, while the red dot a generic point with its generic uncertainty contours. Image adapted from Ref. [84].

This special case is hardly realized in practice, and in a more general situation the relative uncertainty for the two variables decreases differently as the energy increases. However, if the ratio between te uncertainties is *constant* as the energy increases, the slope of the rejection line is related to their magnitude.

To illustrate this point, let us consider a generic pair of variables ξ and ζ, with uncertainties σ_ξ and σ_ζ, satisfying a certain linear relation within each other, and let $\hat{\mathbf{k}}$, $\hat{\mathbf{k}}'$ be respectively the direction vectors for the line representing this linear relation and the cut line

$$\hat{\mathbf{k}} = \frac{1}{\sqrt{1+K^2}}\begin{pmatrix}1\\K\end{pmatrix} \qquad \hat{\mathbf{k}}' = \frac{1}{\sqrt{1+K'^2}}\begin{pmatrix}1\\K'\end{pmatrix}. \tag{5.22}$$

In general, these two vectors have an arbitrary orientation with respect

to each other, though, as stated above, these two lines are orthogonal in the reference frame where σ_ξ and σ_ζ are equal. It is possible to transform (ξ, ζ) into a set of new variables with unitary (equal) variances by means of a whitening transformation (see e.g. [103]). It is evident that the matrix

$$\mathrm{H} = \begin{pmatrix} 1/\sigma_\xi & 0 \\ 0 & 1/\sigma_\zeta \end{pmatrix} \tag{5.23}$$

realizes this transformation. Therefore, in the reference frame where ξ, and ζ have unitary variances, the transformed line vectors $\mathrm{H}\hat{\mathbf{k}}$ and $\mathrm{H}\hat{\mathbf{k}}'$ are orthogonal

$$0 = (\mathrm{H}\hat{\mathbf{k}})^T \cdot (\mathrm{H}\hat{\mathbf{k}}') = \frac{1}{\sigma_\xi^2} + \frac{KK'}{\sigma_\zeta^2} \qquad \Rightarrow \qquad K' = -\frac{\sigma_\zeta^2}{\sigma_\xi^2}\frac{1}{K}, \tag{5.24}$$

which means that the slope of the cut line can be determined by the ratio of the variances of the two variables involved. It follows that, if $\sigma_\zeta/\sigma_\xi =$ constant, for any point the same transformation can be applied with the effect of transforming the 1σ contour into a circle, so that an orthogonal cut line would not introduce any bias (cf. Fig. 5.18-*left*).

If this ratio is not constant, Eq. (5.24) does not hold, because the transformation realized by the matrix in Eq. (5.23) would only be valid for a single data point, or in other terms there would be a different transformation of this type for each point, for which its 1σ contour will appear as a circle. Fig. 5.18 schematically shows the biasing effect as the ratio σ_ζ/σ_ξ is energy dependent, assuming that the relative uncertainties σ_ξ/ξ decrease as the energy increases, while σ_ζ/ζ remains constant. For a point on the calibration line for which the 1σ contour appears as a circle, as one looks to the neighbouring points the 1σ contour becomes an ellipse. At lower energies, Fig. 5.18-*centre*, points below the orthogonal cut line would introduce downwards fluctuations, whilst points above this line would introduce upward fluctuations (see Fig. 5.18-*right*). As the number of data points decrease exponentially with the energy (cf. Fig. 5.17), the contribution to the bias from low-energy data points weights the most. For this reason the slope of the optimal cut line has been deduced by the following Monte Carlo method.

Implementation

In what follows, we present the Monte Carlo simulation used to deduce

the slope of the optimal cut line needed before performing the energy-calibration fit. First, it is necessary to set up the energy conversion for S_{33}, which is known to follow a power law (see e.g. [104])

$$S_{33} = \mathrm{A} \cdot \left(\frac{E_{\mathrm{Infill}}}{E_0}\right)^{\mathrm{B}} \quad \Leftrightarrow \quad E_{\mathrm{Infill}} = E_0 \cdot \left(\frac{S_{33}}{\mathrm{A}}\right)^{1/\mathrm{B}}, \tag{5.25}$$

where the reference energy $E_0 = 0.7 \times 10^{17}$ eV. For the simulation we have chosen A = 17.65 VEM and B = 1.2 (hence *slope* 1/B = 0.83) dimensionless.

As noticed above, the calibration fit needs to be performed above the (Infill) energy threshold. A simple orthogonal cut (w.r.t. the calibration line) in the $(S_{33}, E_{\mathrm{Infill}})$ space would bias the fit, because the relative uncertainties of S_{33} and E_{Infill} scales differently as the air-shower energy increases (cf. Fig. 5.19). One can mimic the deviation from orthogonality of the cut line with respect to the calibration line, defining the cut-line slope δ to be

$$\delta = -(\mathrm{B} + \Delta), \tag{5.26}$$

therefore, the aim of the present simulation is to deduce Δ. With the above definition, if the uncertainties would be equal it will result $\Delta = 0$, implying a cut line orthogonal to the calibration line and hence slope $\delta = -\mathrm{B} = -1.2$. For values $\Delta < 0$ the slope δ decreases and the cut line tends to a horizontal one reproducing the case $\sigma[E_{\mathrm{Infill}}]/\sigma[S_{33}] < 1$. Similarly, for $\Delta > 0$ the cut line tends to a vertical cut, reproducing the case $\sigma[E_{\mathrm{Infill}}]/\sigma[S_{33}] > 1$. The values of Δ were taken in steps of 0.075 in the interval $[-1.5, 1.5]$. For each Δ, the simulation consisted of the following stages, repeated 200 times to collect some statistics:

I. Random extraction of 1000 events from the original energy distribution.
II. The uncertainty to E_{Infill} and S_{33} is assigned according to the specific case: *equal*, *constant*, *realistic* (see below for details).
III. Apply the optimal-line cut for the chosen Δ.
IV. Fit the energy model in Eq. (5.25) to the mock data to estimate the coefficients A and B.

To deduce the optimal Δ value suitable for the real-data calibration, we have repeated the steps above for the three cases of *equal*, *constant*, and *realistic* relative uncertainties:

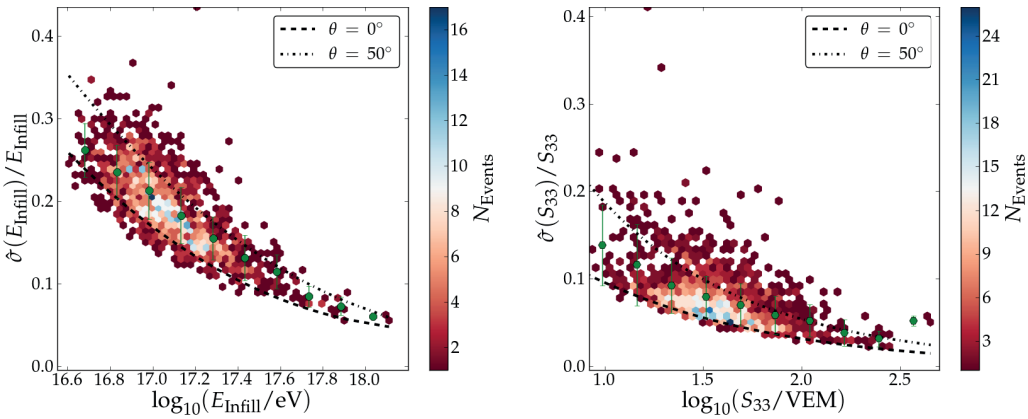

FIGURE 5.19: *Left panel:* Comparison of the model of energy relative uncertainty in Eq. (5.27) against the data. The dashed and the dot-dashed lines represent respectively the model prediction at 0° and at 50°. *Right panel:* Comparison of the model of S_{33} relative uncertainty in Eq. (5.21) against the data. The dashed and dot-dashed lines represent respectively the model predictions at 0° and at 50°.

- *equal uncertainties*: in the first case, uncertainties were drawn according to a normal distribution with mean 10% and spread of 1%, i.e. $\sigma[E_{\rm Infill}]/E_{\rm Infill} = \sigma[S_{33}]/S_{33} = 10 \pm 1\%$.
- *constant uncertainties*: in this second case, uncertainties were taken constant but different, drawing them from normal distributions with $\sigma[E_{\rm Infill}]/E_{\rm Infill} = 5 \pm 1\%$ and $\sigma[S_{33}]/S_{33} = 10 \pm 1\%$.
- *realistic uncertainties*: in this case, the uncertainties for the energy and the reference signal were taken from their respective models. The following (empirical) relative-uncertainty model for the energy estimate, given as:

$$\frac{\hat{\sigma}(E_{\rm Infill})}{E_{\rm Infill}} = a_0 + (a_1 \cos\theta + a_2)\frac{1}{\sqrt{\log_{10} E_{\rm Infill}}} + (a_3 \cos\theta + a_4)\frac{1}{\log_{10} E_{\rm Infill}}, \tag{5.27}$$

and the reference-signal relative-uncertainty model in Eq. (5.21), have been fitted to the data, and are shown in Fig. 5.19. Therefore, for every extracted event $\sigma[E_{\rm Infill}] = \hat{\sigma}(E_{\rm Infill})$ and $\sigma[S_{33}] = \hat{\sigma}(S_{33})$.

Fig. 5.20 shows the results of such a MC simulation, in terms of the estimated calibration parameters for each chosen Δ, and for the specific uncertainty case considered (see legend). In this Figure, the black-dashed horizontal lines are drawn at the original chosen values A = 17.65 [VEM]

(left plot) and B = 1.2 (right plot). Each point in the plot is obtained as the average, and corresponding standard deviation, upon the 200 estimations of the parameters for each $\Delta \in [-1.5, 1.5]$.

As expected, in the equal-uncertainties case the true values for A and B are recovered for $\Delta = 0$, i.e. an orthogonal-line cut would not introduce any bias to the calibration fit, and the reconstructed parameters exactly match the true ones. In the constant-uncertainties case, where $\sigma[E_{\rm Infill}]/\sigma[S_{33}] < 1$, the red triangles in Fig. 5.20 intercept the dashed horizontal line at $\Delta = -0.75$ for the A parameter, while at $\Delta = -0.975$ for the B parameter. This discrepancy is also observed, in the realistic case, where we get two different values for Δ as one considers A or B. For the reconstructed A parameter, we get $\Delta = 1.2$, whereas it seems that the B parameter cannot be accurately estimated, with a systematic deviation of $\simeq 0.03$, which is anyway small compared to the statistical deviation from the average. It is clear that the slope is more sensitive to the upward fluctuations of S_{33}, and that its estimation through the LS-method is not optimal, as the trigger-efficiency cannot be properly taken into account. Finally, a positive value for Δ means that for the realistic case $\sigma[E_{\rm Infill}]/\sigma[S_{33}] > 1$. As it can be noted in Fig. 5.19 as well, the relative uncertainty $\sigma[S_{33}]/S_{33}$ decreases faster with energy than the relative uncertainty $\sigma[E_{\rm Infill}]/E_{\rm Infill}$.

As a result of this MC, for the calibration analysis, an optimal cut line cut with slope $\delta = -(\mathrm{B} + 1.2)$ will be applied before fitting the energy model.

5.3.4 Estimation of the Calibration Parameters

Once the $\Delta = 1.2$ factor for the optimal-line-cut slope has been fixed, we can apply such a cut to the collected data to estimate the energy-calibration parameters. To construct the optimal cut line, a first estimation of the reconstructed constant and slope is needed, as the two lines need to intersect at the energy threshold considered $\tilde{E}_{\rm Infill} \equiv 0.2$ EeV. This preliminary estimate has been obtained by fitting the model in Eq. (5.25) to the data, after rejecting events for which $E_{\rm Infill} < \tilde{E}_{\rm Infill} \equiv 0.2$ EeV. The target function used for this and for all subsequent fits is

$$\sum_{i=1}^{N} \frac{(E_{\rm Infill}[i] - E(S_{33})[i])^2}{\sigma^2_{E_{\rm Infill}}[i] + \left(\frac{\partial E_{\rm Infill}}{\partial S_{33}}\right)^2 \sigma^2_{S_{33}}[i]}. \tag{5.28}$$

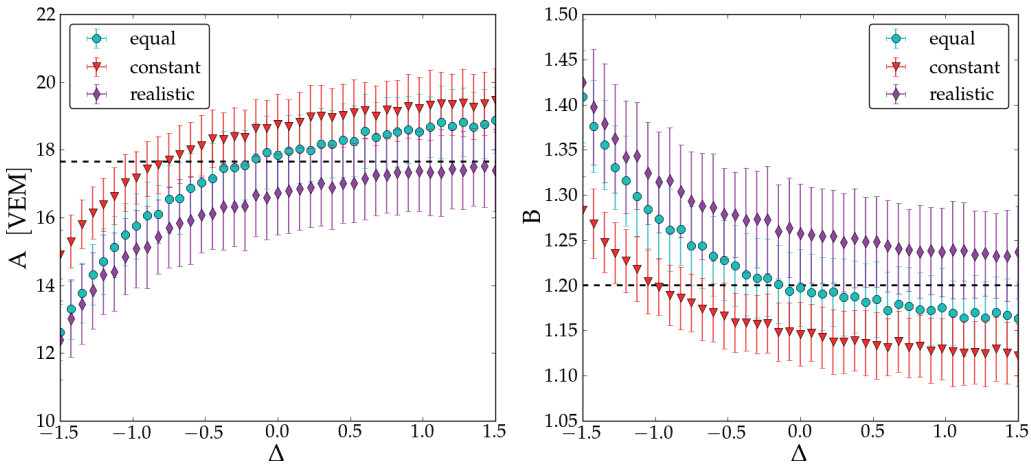

FIGURE 5.20: Result of the Monte Carlo method to infer the optimal Δ value. The data points represent the sample mean and 1σ uncertainty of the 200 estimated values of A and B for each $\Delta \in [-1.5, 1.5]$, and for a specific uncertainty case (see legend). The black-dashed horizontal lines correspond to the initially chosen values $A = 17.65$ VEM, $B = 1.2$ used to convert S_{33} into energy.

As seen before, such a horizontal cut introduces a bias, therefore we opted for a recursive strategy to minimize this unwanted effect. Basically, one uses a first estimate of the parameters to construct the optimal cut line to select events participating in the consecutive fit. In other words, once one has the first estimates of the calibration parameters (A_0, B_0), these parameters are used to construct the minimum-bias optimal cut line with slope $\delta = -(B_0 + 1.2)$, crossing the calibration line at $(\tilde{S}_{33}, \tilde{E}_{\text{Infill}})$, where $\tilde{S}_{33} = A_0 \cdot (E_{\text{Infill}}/E_0)^{B_0}$

Performing the new rejection according to the optimal cut line above, and again the minimization of the target function in Eq. (5.28), a second estimation (A_1, B_1) is obtained, and consequently, a more accurate optimal cut line can be constructed from the latter pair. After few iterations, this process is stopped as the parameter estimates do not change, because at some point, the set of selected data points participating in the minimization does not change. With 186 events participating in the final fit, we obtain the final calibration parameters A, B indicated in Tab. 5.7. In a more natural way, the power-law relation in Eq. (5.25), for the conversion of the reference signal into energy, reads

$$E_{\text{AERAlet}} = \tilde{A} \cdot S_{33}^{\tilde{B}}, \qquad \text{with} \qquad \tilde{A} = E_0 \cdot \left(\frac{1}{A}\right)^{\tilde{B}}, \quad \tilde{B} = \frac{1}{B}, \tag{5.29}$$

which is indeed the relation used in Offline to convert S_{33} to energy. The values of $\tilde{A}, \tilde{B}$ are indicated in Tab. 5.7, with their associated uncertainties obtained by propagating the covariance matrix of A and B.

Parameter	A [VEM]	B	$\tilde{A}$ [PeV/VEM]	$\tilde{B}$
Value	23.16	1.03	3.28	0.974
Error	0.61	0.04	0.18	0.001

TABLE 5.7: The obtained calibration parameters for the conversion of S_{33} into energy.

Fig. 5.21 shows the calibration plot at the last iteration, together with the optimal cut line for $\Delta = 1.2$, and $\Delta = 0$ for comparison. As it can be seen in this Figure, except for the very low-energy events, the calibration line gives a reasonable energy estimate even at lower energies than the threshold $\tilde{E}_{\text{Infill}}$.

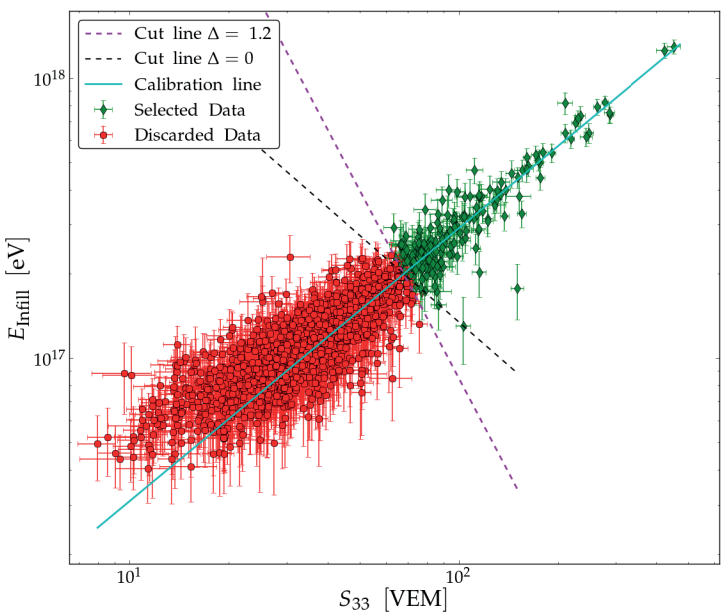

FIGURE 5.21: Energy-calibration plot. The orthogonal cut line with $\Delta = 0$ (black-dashed) has been reported for comparison with the used optimal cut line with $\Delta = 1.2$ (magenta-dashed).

To evaluate the resolution of the energy conversion, Fig. 5.22-*left* shows the histogram of the relative difference $\Delta\bar{E} = (E_{\text{AERAlet}} - E_{\text{Infill}})/\langle E \rangle$ between the converted energy E_{AERAlet}, by making use of Eq. (5.29), and the reconstructed energy E_{Infill} for the selected events. In the same Figure, a Gaussian fit to this histogram is shown, together with its mean μ and

standard deviation σ, which corresponds to an energy resolution for the AERAlet energy estimation of about 21%.

Fig. 5.22-*right* shows this relative difference as a function of $E_{\rm Infill}$, for each event included in the fit. Although two outliers are visible in the top-left corners, with large $E_{\rm Infill}$ uncertainty, no appreciable drift is observed as a function of energy above the threshold. The same comparison for the whole dataset, hence including events below the threshold, is shown in Fig. 5.23. In this Figure, the two-dimensional distribution of the relative difference versus the Infill energy is plotted together with a binned profile in $\log_{10} E_{\rm Infill}/{\rm eV}$, to indicate the average deviation. The average value of $\Delta\bar{E}$ in the first bin, for the selected events (in green), is appreciably above the rest of the bins, as $\Delta\bar{E}$ is appreciably below in the last bin for the discarded events (in red). This effect is due to the regular binning in $\log_{10} E_{\rm Infill}/{\rm eV}$ careless of the applied tilted cut, as it is visible in Fig. 5.22-*right* if one imagines to average on the first few points from the left. It is interesting to notice that the average $\Delta\bar{E}$ for the rejected events is systematically < 0, which is a sign of the upward fluctuations caused by the trigger efficiency of the Infill well below the threshold, while it fluctuates around 0 for the selected events.

At the end of Section 5.3.1 we found that for events below $S_{33}^{\rm thr} \equiv 29.85$ [VEM], the $\cos^2(\theta)$ distribution is no longer uniform, or in other words the assumption of an isotropic flux does not hold because of the detector efficiency. Hence, one can consider $S_{33}^{\rm thr}$ as the trigger threshold for the AERAlet array. It follows that, by means of the energy calibration established, the estimated AERAlet energy threshold is $\simeq 0.09$ EeV $\equiv \tilde{E}_{\rm AERAlet}$.

5.4 Low-Energy Extension of the Energy Spectrum

This Section presents the calculation of the energy spectrum for the CR events observed with the AERAlet facility at the Pierre Auger Observatory from May 2013 to February 2015.

The natural step after the energy calibration is to exploit the energy estimate for AERAlet to calculate the flux of CRs below the Infill threshold. For this aim, the energy-conversion coefficients have been then inserted into the $\overline{\rm Off}\underline{\rm line}$ configuration, and the set of raw data used earlier is again reconstructed. The flux of cosmic rays J as a function of energy is given

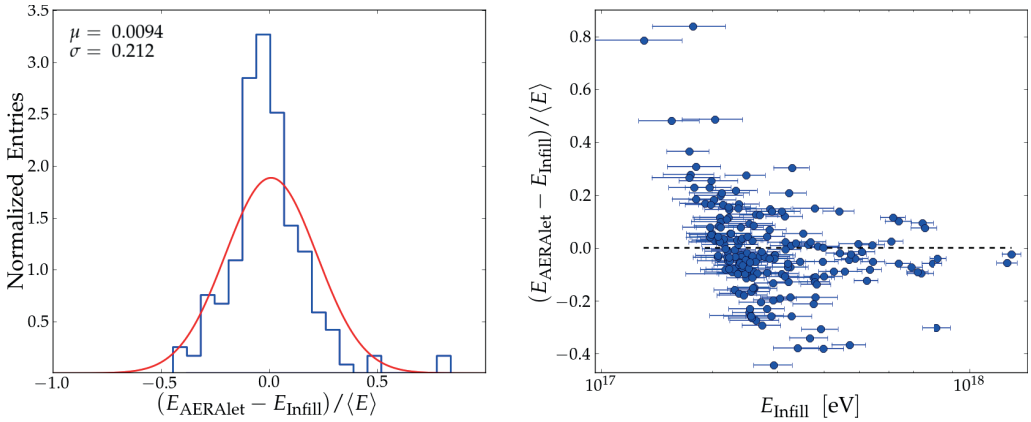

FIGURE 5.22: *Left panel:* Distribution of the relative difference between the converted energy E_{AERAlet} and the reference energy E_{Infill}: $(E_{\text{AERAlet}} - E_{\text{Infill}})/\langle E\rangle$. The energy resolution is given by the 1σ spread of this distribution indicated in the top-left corner. *Right panel:* Scatter plot of this relative difference as a function of E_{Infill}. As expected the spread is large for low-energy events.

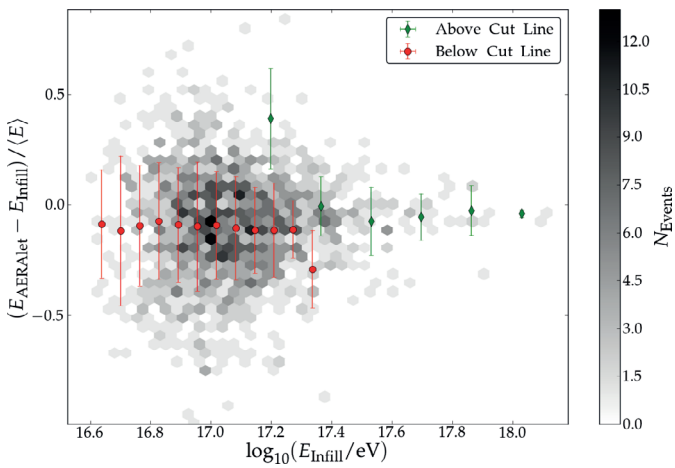

FIGURE 5.23: Two-dimensional distribution of $\Delta\bar{E} = (E_{\text{AERAlet}} - E_{\text{Infill}})/\langle E\rangle$ versus $\log_{10} E_{\text{Infill}}$ for selected and rejected events (see legend), with a superimposed binned profile indicating the average and 1σ spread of this difference in each $\log_{10} E_{\text{Infill}}$ bin. For events below the threshold the average difference appreciably deviates from zero, whereas it oscillates around zero above the threshold.

as in Eq. (5.10), or operationally defined as:

$$J(E) =: \frac{d^2N}{dE\,d\Sigma} \simeq \frac{\Delta N(E)}{\Delta E}\frac{1}{\Sigma}, \tag{5.30}$$

where N is the number of events with energy between E and $E + dE$,

and $d\Sigma = dA_{\rm eff}\, d\Omega\, dt$ is the differential exposure, in which the effective observation time and area are combined. In practice, the flux can be approximated by counting the number of events $\Delta N(E)$ in the energy bin ΔE, and dividing by the integrated exposure Σ, which is constant only above the energy threshold. By dividing by the exposure, essentially, all the detector-dependent quantities are divided out, and the calculated flux can be directly compared with the one obtained by other experiments.

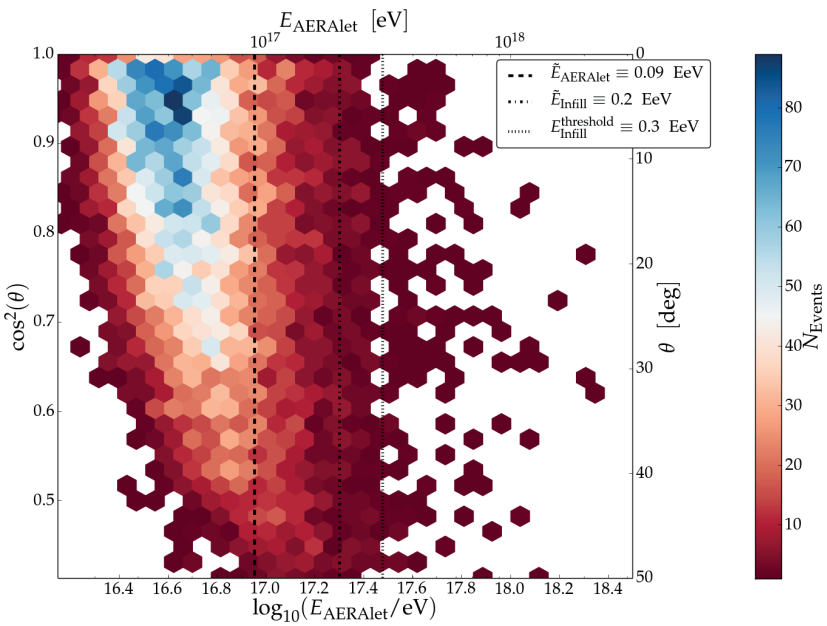

FIGURE 5.24: Energy–$\cos^2(\theta)$ event distribution for AERAlet-only reconstructed events passing the 6T5 requirement, see Section 4.2. (the corresponding zenith angles are indicated in degrees by the right scale).

The two dimensional zenith-angle/energy distribution of the 10187 selected 6T5 events, with $\theta < 50^\circ$, is shown in Fig. 5.24. The dashed line in the plot indicates the energy-threshold estimate $\tilde{E}_{\rm AERAlet}$. In the same figure, the dot-dashed and the dotted lines respectively indicate, as in Fig. 5.17, the infill threshold at 0.3 EeV and its less strict value at 0.2 EeV. It is evident that below $\tilde{E}_{\rm AERAlet}$, the atmospheric attenuation of the electromagnetic shower component plays a major role in preventing the detection of low-energy showers. However, in this energy regime, one can identify a zenith-angle energy-dependent threshold $\theta_{\rm thr}(\Delta E)$ for each energy bin ΔE, below which the $\cos^2\theta$ event distribution is still uniform.

As the integrated exposure Σ depends on the subtended solid angle, below $\tilde{E}_{\rm AERAlet}$, it is possible to determine $\theta_{\rm thr}(\Delta E)$ to provide a heuristic

correction to the calculated exposure. This correction is meant to be able to use as many events as possible, though it can only be applied for energies such that $\log(E/\mathrm{eV}) \gtrsim 16.6$. Around this energy, a drastic drop is observed due to the vanishing trigger efficiency, and the use of events below this limit would require extensive simulation studies to estimate an energy-dependent efficiency-correction for the integrated exposure.

5.4.1 Exposure Calculation

The exposure is defined as the time integrated *aperture*, which is given by the effective detector area integrated over the subtended solid angle. In turn, the effective area coincides with the geometrical one as the shower core is well contained within the boundary of the array, which is guaranteed by the T5 condition (see Section 4.2). At the same time, the acceptance of the detector needs not to depend on the nature of the primary particle, its energy or arrival direction, which is true at full efficiency conditions in the chosen zenith-angle range. As reported in [105], a simulation-based study shows that the 433 m array guarantees a 95% detector efficiency at energies in excess of $\approx 9 \cdot 10^{16}$ eV for proton and $6 \cdot 10^{16}$ eV for iron primaries.

Fig. 5.25 shows the detector efficiency obtained in [105], as a function of energy, where a detection is counted for having a (positive) trigger in at least three detectors. The trigger probability is a function of the observed signal strength, and therefore of the energy of the CR event. This probability has been estimated using simulated proton- and iron-initiated air showers, as the ratio $N_{\mathrm{obs}}/N_{\mathrm{exp}}(E) := \epsilon(E)$ of the number of events giving rise to a positive trigger, hence being *observed*, and the number of generated (*expected*) events.

Using this result, to calculate the spectrum in Section 5.4.3, we shall estimate the number of expected events in a given energy bin as $\hat{N}(E_k) \approx N(E_k)/\epsilon(E_k)$. For this purpose, the function $\hat{\epsilon}(E)$ below has been fitted to the data from [105] in order to be able to retrieve the efficiency value at the desired energy:

$$\hat{\epsilon}(E) = \frac{a_0 \cdot \mathrm{e}^{a_2 a_3} + a_1 \cdot \mathrm{e}^{a_3 \log(E/\mathrm{eV})}}{\mathrm{e}^{a_2 a_3} + \mathrm{e}^{a_3 \log(E/\mathrm{eV})}}. \tag{5.31}$$

As one can see in Fig. 5.25, the fitted curves $\hat{\epsilon}^{\mathrm{p}}(E)$ (solid line), and $\hat{\epsilon}^{\mathrm{Fe}}(E)$ (dashed line) are slightly shifted with respect to the data points, which results in 5% positive deviation in the number of expected events. This

shift has been considered as a systematic uncertainty in the number of events, which is then included in the calculation of the flux uncertainty.

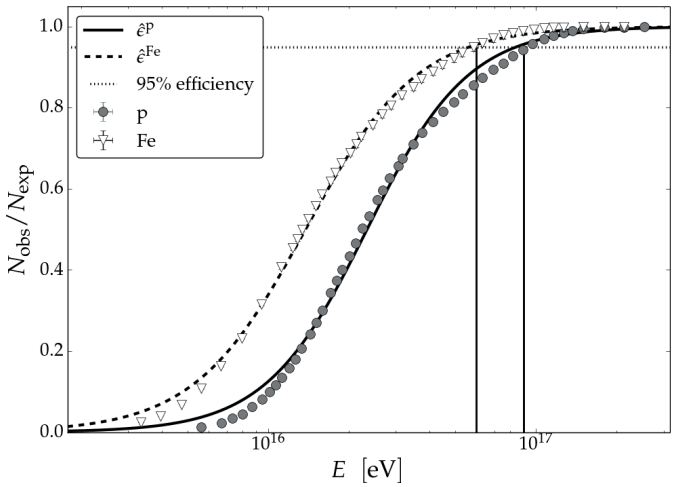

FIGURE 5.25: Relative acceptance for AERAlet from [105] as a function of the primary energy, for protons (gray circles) and irons (white triangles).

As mentioned above, the exposure $d\Sigma = dA_{\text{eff}}\, d\Omega\, dt$ summarizes the observational detector capabilities combining the effective sensitive area dA_{eff}, the solid angle $d\Omega$, and the effective time dt.

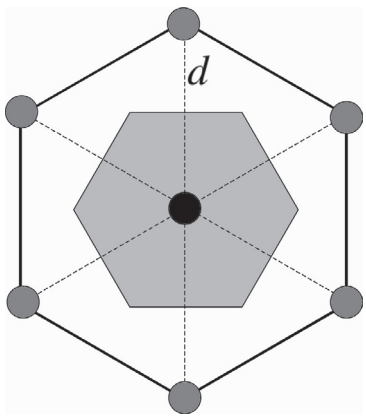

FIGURE 5.26: Simple sketch of the AERAlet elementary observation area, with d the array spacing. Image adapted from Ref. [106].

The effective sensitive area is obtained taking into account the zenith angle, by projecting the elementary surface element onto the shower plane $dA_{\text{eff}} = A_{\bigcirc} \cos\theta\, d\Omega$. The elementary surface element $A_{\bigcirc}$ is nothing but the elementary unit of the array capable of recording and reconstructing air

showers, i.e. the hexagonal cell made up of seven active stations, a central one with six neighbours. Therefore, the 6T5 condition greatly simplifies the exposure calculation, as $A_{\hexagon}$ is constant, and the time integral is reduced to the sum of the seconds when all stations of the array were active.

As seen in Section 4.2, this condition is implemented by the posterior-T5 condition requiring 6 working stations around the one with the largest signal (6T5). Thanks to this condition, the calculation of the elementary-cell area is straightforward, as it can be simply obtained as the Wigner–Seitz cell area (cf. Fig. 5.26) around a lattice point (the central tank):

$$A_{\hexagon} = \frac{\sqrt{3}}{2} d^2 \simeq 0.16 \text{ km}^2. \tag{5.32}$$

Here, d is the array spacing for AERAlet, which has been taken as the average distance (429.5 m) of the actual station distances forming the unique hexagon. As mentioned above, the effective sensitive area A_{eff} takes into account the subtended solid angle, and ultimately depends on the zenith-angle threshold:

$$\begin{aligned} A_{\text{eff}}(\theta_{\text{thr}}) &= \int dA_{\text{eff}} = A_{\hexagon} \int \cos\theta d\Omega = A_{\hexagon} \int \cos\theta \, \sin\theta \, d\theta \, d\phi \\ &= A_{\hexagon} \int_0^{2\pi} d\phi \int_{\cos\theta_{\text{min}}}^{\cos\theta_{\text{thr}}} \cos\theta \, (-d\cos\theta) = A_{\hexagon} \cdot \pi \cdot (\cos^2\theta_{\text{min}} - \cos^2\theta_{\text{thr}}) \\ &= A_{\hexagon} \cdot \pi \cdot (1 - \cos^2\theta_{\text{thr}}), \end{aligned} \tag{5.33}$$

where $\theta_{\text{min}} = 0^\circ$, and $\theta_{\text{thr}} = 50^\circ$ for events with $E > \tilde{E}_{\text{AERAlet}}$ or $\theta_{\text{thr}} = \theta_{\text{thr}}(E)$ for events with $E < \tilde{E}_{\text{AERAlet}}$. Above the threshold, one has $A_{\text{eff}}(50^\circ) \simeq 0.29 \text{ km}^2 \cdot \text{sr}$, and what is left for the exposure calculation is the integral over the active observational time. In turn, this integral is reduced to the sum of the time intervals (in seconds) during which all seven stations of the hexagon were functioning (cf. Fig. 5.27):

$$\Sigma = \int_{t_0}^{t_1} A_{\text{eff}}(50^\circ) \, dt = A_{\text{eff}}(50^\circ) \sum_k \Delta t_k = 0.44 \text{ km}^2 \cdot \text{yrs} \cdot \text{sr} \tag{5.34}$$

As mentioned in Section 5.1, T2 files are used to estimate the time intervals Δt_k, as these files store the detector activity second by second. In addition, the method described in [107], based on the daily rate of T5 events, has been used to check the presence of unstable data-taking periods related to T3 errors. Such errors eventually occur at the data-

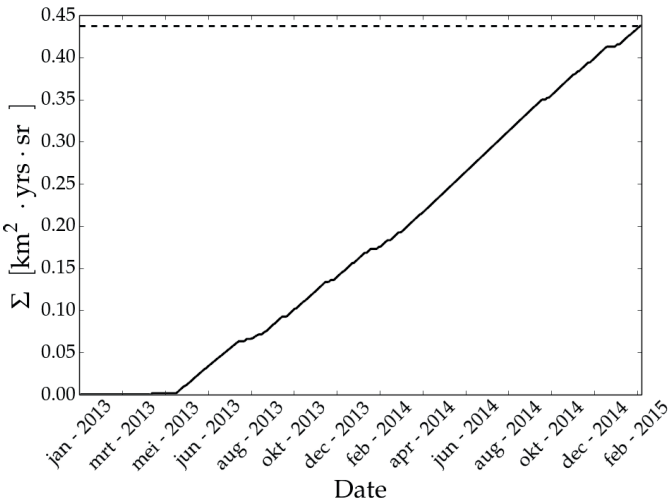

FIGURE 5.27: The calculated exposure for the single hexagon of the 433 m SD infill array (AERAlet) as a function of time, based on the T2 monitoring files.

acquisition level, and are not related to the single-station activity. Thanks to this method, several days have been identified during which the uptime was zero, although T5 events were present and successfully reconstructed. One can conclude that for these days the T2 information was either corrupted or completely missing. As it is not possible to know the uptime without T2-file information, events occurring during these days have been discarded.

Fig. 5.28 shows the daily-rate empirical distribution for the selected T5 events, together with a Gaussian fit to it (red-solid line). It should be noted that there are days with an extremely high rate of T5 events, outside the 99% confidence interval of the fitted density (more than 37 events per day). Again, during these days the T2 trigger-rate monitoring likely had problems. For this reason, events occurring in these days have been rejected as well, and the exposure was corrected accordingly ($\approx$ 14.5 hrs in total). In this same Figure, the black-solid line is the fitted density after the exclusion of these four days.

After these additional cuts, the total number of events has dropped to 9782.

5.4.2 Heuristic Correction

The exposure calculation in Eq. (5.34) is valid only above the energy threshold $\tilde{E}_{\text{AERAlet}}$ where attenuation effects do not limit the flux of ac-

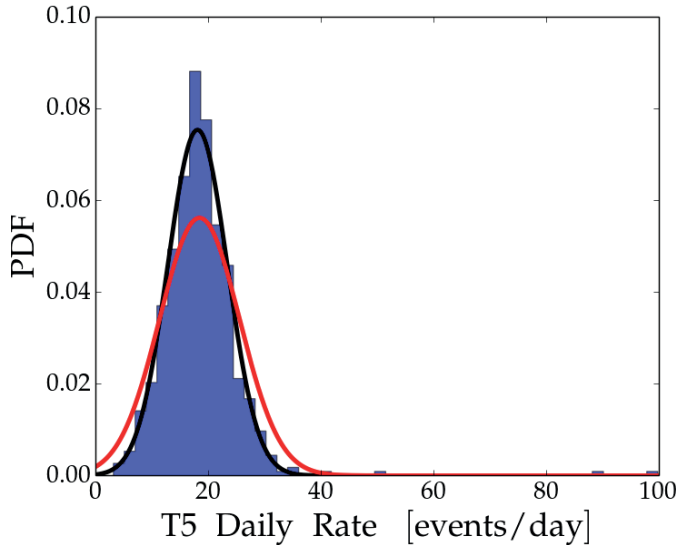

FIGURE 5.28: Distribution of the daily rate for T5 events observed by AERAlet. The red- and black-solid curves are Gaussian fits to this distribution respectively before and after removing the anomalous days of extremely-high rate.

tual CR events being observed. Below this threshold, it is necessary to take into account the zenith-angle dependency outlined in Eq. (5.33). At the same time, as pointed out before, this correction can only be applied above a certain energy, i.e. $\log(E_{\rm AERAlet}/{\rm eV}) \simeq 16.6$, below which a drastic drop is observed as well for nearly vertical events ($\theta \lesssim 15^\circ$). In other words, we are assuming that between this limit and the $\tilde{E}_{\rm AERAlet}$ threshold, the number of detected events mainly drops because the atmospheric attenuation. Hence, a zenith angle limit can be found for which the energy event distribution is consistent with a uniform distribution. Combining Equations (5.34) and (5.33), one has

$$\Sigma(\theta_{\rm thr}) = A_{\rm eff}(\theta_{\rm thr}) \sum_k \Delta t_k = A_{\rm eff}(\theta_{\rm thr}) \cdot T, \tag{5.35}$$

where $T = \sum_k \Delta t_k$, and $\theta_{\rm thr}$ this desired limit. As a matter of fact $\theta_{\rm thr} = \theta_{\rm thr}(E_i)$ where E_i is the central value of the i-th bin of width ΔE_i. The energy bins are defined such that $\log_{10}(\Delta E_i/{\rm eV}) = 0.1 \;\; \forall i$. In each bin, the value of $\theta_{\rm thr}(E_i)$ is deduced empirically as the largest angle for which the $\cos^2\theta$ distribution of events such that $\theta < \theta_{\rm thr}$ is compatible with a uniform distribution.

To find $\theta_{\rm thr}(E_i)$, for each energy bin, we adopted the following strategy:

(i) Starting from $\theta_{\text{thr}} = \theta_{\text{init}} \equiv 49.9°$, we select a sample of events with $\theta < \theta_{\text{thr}}$.

(ii) The $\cos^2\theta$ distribution of this selected sample is constructed from the data with a variable number of bins $N_{\text{bins}} \in \{10, 11, \ldots, \lfloor N_{\text{samples}}/3 \rceil\}$ (with $\lfloor N_{\text{samples}}/3 \rceil$ the closest integer to the number of samples divided by three).

(iii) For each value of N_{bins}, a constant is fitted to the entries n_i with weights $\sqrt{n_i}$ $(i = 1, 2, \ldots, N_{\text{bins}})$, and the corresponding χ^2_{red} is collected.

(iv) If the median χ^2_{red} of this collected statistics is between 0.9 and 1.2, then the current value of θ_{thr} is assumed as zenith-angle threshold (this is done as the χ^2_{red} value might depend on the number of bins in which one splits the data).

(v) Otherwise, the value of θ_{thr} is decreased, and the steps (ii) – (iv) are repeated.

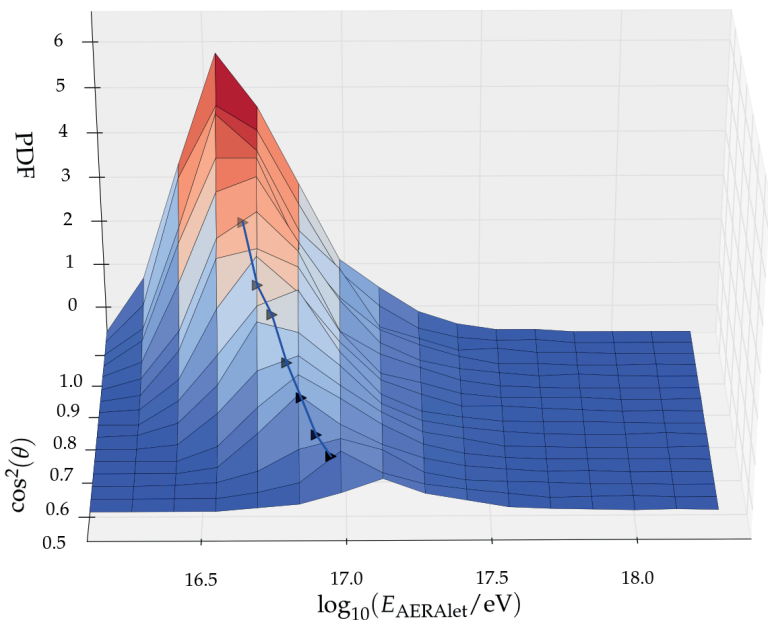

FIGURE 5.29: The calculated zenith-angle (efficiency) thresholds (black triangles), superimposed to the $\cos^2(\theta)$–energy event distribution. The binning of the empirical distribution is not related to the considered energy bins used for the calculated angles. The blue line has been drawn to guide the eye.

Fig. 5.29 shows the result of such a procedure, superimposed on the $\cos^2\theta$-energy distribution of the used CR events, where the coloured con-

tours are meant to guide the eye in visualizing the event density. The black triangles indicate the determined values of $\theta_{\rm thr}(E_i)$. The associated uncertainty $\sigma_{\theta_{\rm thr}}(E_i)$ of these thresholds are in the order of few degrees, though they are not shown in the Figure for clarity reasons. The source of this uncertainty is mainly related to the energy determination and partially to the zenith-angle uncertainty. Indeed, the energy is determined by Eq. (5.29), where the calibration parameters are in turn affected by their own uncertainties. As the data are separated into energy bins, close to the bin edges, a variation of the energy value has the effect of changing the population in these bins, and therefore the $\cos^2\theta$ distribution will be affected as well.

Taking into account this fact, the uncertainty $\sigma_{\theta_{\rm thr}}(E_i)$ has been determined by repeating 200 times the procedure above. Each time the energy-calibration parameters are extracted from a binomial normal distribution constructed from their fitted values and associated covariance matrix. Thereafter, the energy is determined according to these varied parameters, and the data separated in energy bins. Hence, repeating the steps (i) – (v) above, for each energy bin, one obtains a distribution of $\theta_{\rm thr}(E_i)$ values from which the sample variance can be calculated.

Once the zenith-angle thresholds have been found, Eq. (5.35) will be used to provide the corrected exposure for the CR flux calculation.

5.4.3 AERAlet Energy Spectrum

As defined by Eq. (5.30), the energy spectrum is calculated by counting the number of CR events whose energy lies in a certain bin ΔE_k. The binning in energy is performed on its logarithm, with a bin width of 0.1 up to $\log_{10}(E_k/\mathrm{eV}) = 17.35$, and of 0.2 afterwards to accommodate for the low event number. Then, the bin counts are respectively divided by the energy interval and by the exposure, respectively in Eq. (5.34) and Eq. (5.35), above and below the energy threshold.

As mentioned in Section 5.4.1, below the energy threshold, the number of detected events underestimates the number of actual events that would have been detected with a fully-efficient detector. One can take care of this deficiency by exploiting the efficiency curve shown in Fig. 5.25, which depends on the assumed composition. Therefore, for each k-th energy bin, together with the data counts, one can estimate the number of events $\hat{N}_k^{\rm p,Fe}$ that one would have been observed if all incoming CRs

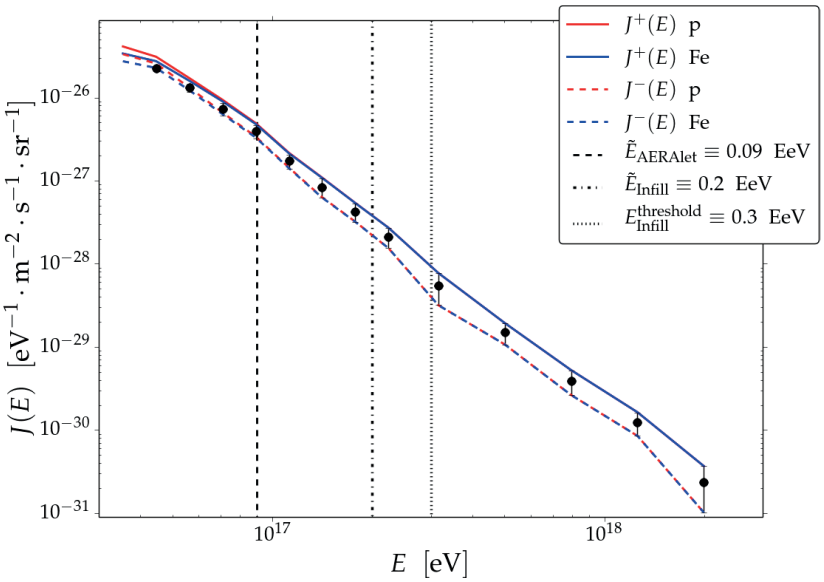

FIGURE 5.30: The energy spectrum obtained by AERAlet measurements (black dots). The red (blue) dashed and solid lines provide the $\pm 1\sigma$ uncertainty for proton (iron) assumed composition as in Eq. (5.39). The vertical lines indicate the energy thresholds for AERAlet (dashed) and Infill in its soft (dot-dashed) and strict (dotted) versions.

would have been protons (irons) as:

$$\hat{N}_k^{\rm p,Fe} \simeq \frac{N(E_k)}{\hat{\epsilon}^{\rm p,Fe}(E_k)}. \tag{5.36}$$

Here, $\hat{\epsilon}^{\rm p,Fe}$ is the fraction of events satisfying the trigger condition, defined by Eq. (5.31) and shown as the curve in Fig. 5.25. Hence, according to Eq. (5.30), for the flux-spectrum corrected values one has

$$J^{\rm p,Fe}(E_k) = \frac{\Delta \hat{N}_k^{\rm p,Fe}}{\Delta E_k \, \Sigma(\theta_{\rm thr})}, \tag{5.37}$$

in the energy bin ΔE_k, and with $\Sigma(\theta_{\rm thr})$ defined in Eq. (5.35).

Fig. 5.30 shows the result of such a procedure. The black data points indicate the calculated energy spectrum from the data below and above the adopted energy threshold for AERAlet, indicated by the black-dashed line. For comparison, the dotted line indicates the Infill threshold at 0.3 EeV, while the dash-dotted line represents its less strict value at 0.2 EeV. The solid- and dashed-red (blue) lines respectively indicate the expected flux values as in Eq. (5.37) assuming a pure proton (iron) composition

increased (J^+) and decreased (J^-) by the 1σ uncertainty.

The uncertainties $\sigma[J(E_k)]$, of the binned flux values $J(E_k)$, result from the propagation of the energy-estimation uncertainty, which in turn depends on the uncertainties of energy-conversion parameters found in the previous Section and specified in Tab. 5.7. Indeed, the number of events in the k-th bin is ultimately determined by the converted energy value by means of Eq. (5.29), which in turn inherits the shower-to-shower and sampling fluctuations of the estimator. As a result, upward and downward fluctuations do affect the bin population, as the distribution of events in a certain bin changes according to the final energy estimate.

In addition, another source of uncertainty is introduced by the zenith-angle-threshold estimate used in Eq. (5.35), which carries its own uncertainty. Though it influences the spectrum values to a lesser extent than the energy determination, its effect has been included in the total uncertainty by quadrature:

$$\sigma^2[J(E_k)] = \sigma^2_E[J(E_k)] + \sigma^2_{\theta_{\rm thr}}[J(E_k)]. \tag{5.38}$$

Here, the variances $\sigma^2_E[J(E_k)]$ and $\sigma^2_{\theta_{\rm thr}}[J(E_k)]$, respectively due to the propagated energy and zenith-angle-threshold uncertainties, have been estimated following the steps listed below:

1. A random realization of the energy-calibration parameters $(\mathrm{A}^{(s)}, \mathrm{B}^{(s)})$ is extracted from the bivariate normal probability density function (p.d.f.) $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} \equiv (\mathrm{A}, \mathrm{B})$, and $\boldsymbol{\Sigma}$ their covariance matrix, resulting from the minimization of Eq. (5.28).
2. The varied energy $E^{(s)}$ is determined using Eq. (5.29), according to the varied energy-calibration parameters, for all selected events.
3. The varied flux values $J_k^{(s)} = J(E_k^{(s)})$ are obtained as above, considering the varied energies $E^{(s)}$.
4. The steps (1.) – (3.) are repeated $S = 200$ times, in order to obtain for each bin a distribution of flux values $\{J_k^{(s)}\}_{s=1}^{S}$.
5. Finally, from this distribution, one can estimate the sample variance $\sigma^2_E[J(E_k)]$.
6. Similarly, a random realization $\theta_{\rm thr}^{(l)}$ is extracted from $\mathcal{N}(\theta_{\rm thr}, \sigma_{\theta_{\rm thr}})$, for each energy bin. Then again, the varied spectrum values $J_k^{(l)} = J(\theta_{\rm thr}^{(l)}(E_k))$ are calculated using the varied zenith-angle thresholds. For each bin k, repeating these steps $L = 200$ times, we obtain a

distribution $\{J_k^{(l)}\}_{l=1}^{L}$ from which we estimate the sample variance $\sigma^2_{\theta_{\rm thr}}[J(E_k)]$.

In addition, a systematic contribution to the uncertainty of the pure proton/iron spectra arises from the difference between the fitted curve $\hat{\epsilon}(E)$ in Eq. (5.31) and the efficiency values $\epsilon(E_k)$, visible in Fig. 5.25. As a consequence, this discrepancy gives rise to a different number of events when making use of Eq. (5.36) to estimate the efficiency correction for the spectrum. This systematic contribution is added quadratically to the uncertainty estimated by Eq. (5.38). Finally, the $\pm 1\sigma$ deviation $J^{\pm}(E)_{\rm p,Fe}$ shown in Fig. 5.30 are given by

$$J^{\pm}(E)_{\rm p,Fe} = J^{\rm p,Fe}(E) \pm \sqrt{\sigma^2[J^{\rm p,Fe}(E)] + \sigma^2_{\rm sys}[J^{\rm p,Fe}(E)]} \tag{5.39}$$

As mentioned in Section 2.1, the energy spectrum follows a power law $J(E) \propto E^{-\gamma}$, with γ known as the spectral index. Its steep fall implies that low-energy bins are more populated than high-energy bins, which in turn means that the flux relative uncertainties get large as the energy increases. Therefore, the flux values at the lowest energy, namely the ones corrected assuming a light/heavy composition, will have the largest weight on the spectral-index fitting. At these energies, it can be noted in Fig. 5.30 that the uncorrected flux values (black dots) fall on top of the -1σ dashed line for protons (and even below for the first point). Hence, to avoid biasing the fit of the spectral index, one can consider to take into account the corrected data points, assuming a pure-proton or a pure-iron composition.

A global fit of a power law can be alternatively performed in a linear fashion in the log-log representation, minimizing the target function

$$\sum_{k=1}^{N_{\rm bins}} \frac{\{\log \tilde{J}(E_k) - (\kappa - \gamma \log \tilde{E}_k)\}^2}{\sigma^2[\log \tilde{J}(E_k)]}, \tag{5.40}$$

where $\tilde{J}$, $\tilde{E}_k$ are the dimensionless versions of J and E_k, and $\sigma^2[\log \tilde{J}(E_k)]$ obtained by analytical error propagation. The minimization above returns, for the binned spectra obtained by means of Eq. (5.37), $\gamma_{\rm p} = 3.06 \pm 0.13$, for proton-assumed composition, and $\gamma_{\rm Fe} = 3.01 \pm 0.13$ for iron-assumed composition. In both cases, the minimum reduced χ^2 is very low, because of the large uncertainties of the flux values, which explains the large uncertainty on γ. This is mainly due to the very limited energy resolution of 21%, which propagates down to the flux estimation, further amplified by its steep fall.

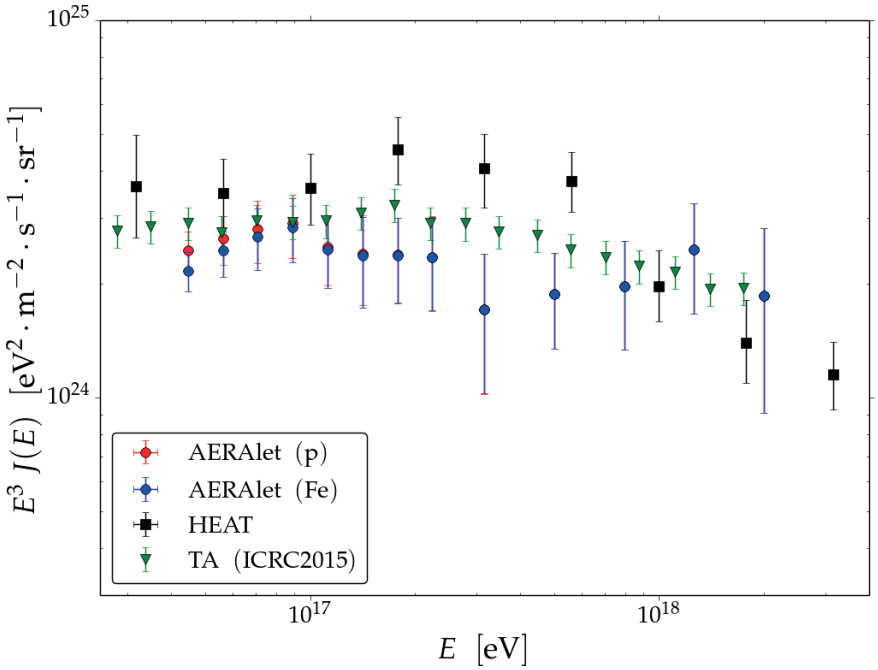

FIGURE 5.31: The energy spectrum obtained by AERAlet measurements, scaled with E^3, assuming a light (heavy) composition to correct for the efficiency just below the threshold in red (blue). The comparison is with the scaled versions of the HEAT (black squares) and TA (green triangles) spectra.

As a consequence, the estimated spectral index over the whole energy range is $\sim 2\sigma$ away with respect to the spectral index provided by other experiments such as HEAT, and Telescope Array (TA). The first report a spectral index of -3.23 [73] and the second of -3.30 [108], in approximately the same range of energies. Evidently, the very low reduced χ^2, resulting from the fit of the spectral indexes, indicates that to perform such a comparison one would need more data in order to reduce the flux uncertainties.

Nevertheless, one can compare the AERAlet spectrum with the spectra obtained by HEAT (from [73]) and TA (from [109]), scaling all of them with E^3 to improve the visualization of the features. Fig. 5.31 shows this comparison with the HEAT spectrum in black squares and the TA spectrum in green triangles. As the values for TA have been extracted digitizing[1] the plot, a 10% uncertainty has been assigned. Again, for AERAlet, we plot the corrected values of the flux assuming a light (red dots) and heavy (blue dots) composition.

[1]For this purpose we used the public `WebPlotDigitizer` software [110].

Although the average energy-spectrum slope only agrees at 2σ level, the AERAlet spectrum shows a very good agreement with the TA spectrum up to the energy threshold at $\sim 10^{17}$ eV. After that, though the spectra are consistent at the 1σ level, we observe what looks like a drop in the number of detected events. Around this energy, the KArlsruhe Shower Core and Array DEtector (KASCADE) – Grande experiment, have found a steepening of the spectrum at $10^{16.92}$ eV for the heavy CR component [111], and a softening of the spectrum at $10^{17.08}$ eV for the light CR component [112]. Another drop in the number of detected event is observed at $\approx 3 \cdot 10^{17}$ eV. Around this energy, an additional steepening of the spectral index was identified in [87], and at a slightly higher energy in [73].

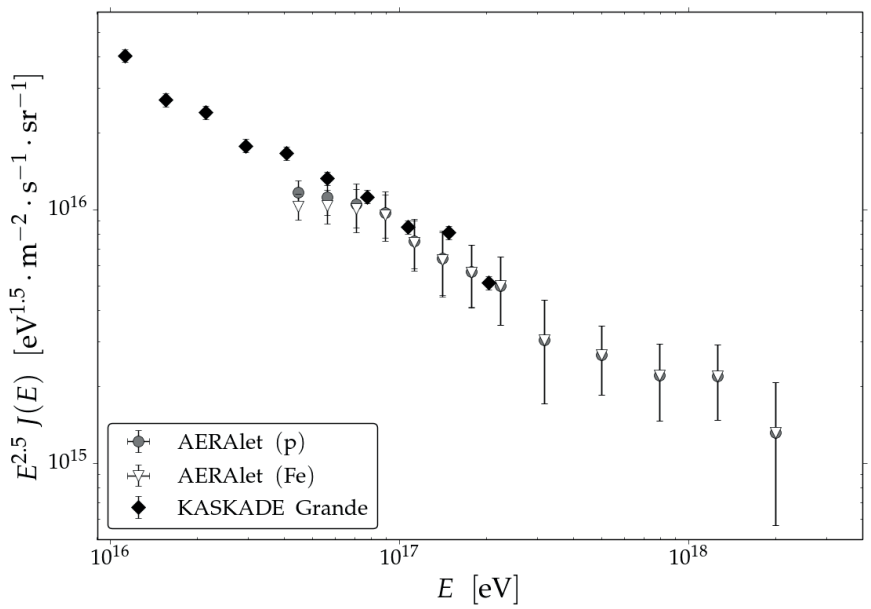

FIGURE 5.32: The energy spectrum obtained by AERAlet measurements, scaled with $E^{2.5}$, assuming a light (heavy) composition to correct for the efficiency just below the threshold, in gray dots (white tringles). The AERAlet spectrum is compared with the scaled version of the KASCADE – Grande all-particle spectrum from Ref. [113]

Fig. 5.32 shows the comparison of the AERAlet spectrum with the KASCADE – Grande all-particle spectrum extracted from Ref. [113]. Again, for AERalet we represent the corrected values assuming pure-proton (gray circles) and pure-iron (white triangles) compositions. The spectra show a very good agreement in the overlapping region, and the AERAlet spectrum conveys the impression of being a natural continuation of the KASCADE – Grande all-particle spectrum. The latter has been deduced digi-

tizing the plot, a 5% uncertainty is added for comparison. Nevertheless, the last two data points in [113] present a larger uncertainty than this 5%, mainly due to the low event statistics at the highest energies. On the contrary for the AERAlet spectrum, the lowest energy bins are affected by the low efficiency, and although a correction has been applied, a discrepancy is observed with the KASCADE – Grande spectrum.

5.5 Conclusions

This Chapter focused on EAS measurements performed by means of the AERAlet low-energy enhancement at Pierre Auger Observatory, a 433 m SD infill array. Because these detectors have recently been deployed, we went through the whole process of understanding the detector and its reconstruction performance, with the ultimate goal of obtaining an estimate of the primary CR energy at energies less than $3 \cdot 10^{17}$ eV. This step has mainly consisted in optimizing the LDF parameters for the air-shower reconstruction in Offline, which in fact depend on the array spacing, and consequently in finding the suitable energy estimator. The primary energy estimate has been obtained by calibrating the estimator with the 750 m Infill energy measurement.

With such narrow-spaced stations (with respect to the other arrays at the Observatory), we were able to measure air showers down to $\approx 5 \cdot 10^{16}$ eV, though at these energies, atmospheric-attenuation effects appreciably limit the observational solid angle. On the one hand, this limitation makes the calculation of the detector exposure more difficult, on the other hand, muon-rich showers have higher probability of being detected than muon-poor ones. To overcome this problem, we used a heuristic approach to determine in a given energy bin the maximum zenith angle for which the distribution of CR events is consistent with a uniform distribution. In this way, we have limited the solid-angle integration for the exposure to these energy-dependent zenith-angle thresholds.

Although more detailed simulation studies of the zenith-angle and energy dependencies of the detector efficiency are desirable, this approach has given us the opportunity of exploiting the data below the estimated energy threshold for AERAlet. As a matter of fact, we found a very good agreement with the energy spectrum measured by the TA experiment also in this energy range. And unfortunately, the effective data-taking period was less than expected due to initial (unavoidable) technical problems. Nevertheless, the observed CR flux is consistent within 1σ with the

CR flux measured by other experiments, and though it was not possible to further investigate the spectral features, in correspondence of those we see drastic drops in the number of detected events. Finally this means that, despite the limited energy resolution, the achieved energy calibration can be used for further low-energy CR measurements at the Observatory.

"There is one feature I notice that is generally missing in cargo cult science *... It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty – a kind of leaning over backwards.*

For example, if you're doing an experiment, you should report everything that you think might make it invalid – not only what you think is right about it; other causes that could possibly explain your results; and things you thought of that you've eliminated by some other experiment, and how they worked – to make sure the other fellow can tell they have been eliminated."

R. P. Feynman

6 Measurement Sensitivity of the Radially-Polarized Radio Emission Component

The present Chapter focuses on AERA-I measurements of the radio emission from EASs, detected in coincidence with the AERAlet SD array. The lateral dependence of the electric-field *polarization* gives a very clear signal for the development of the electromagnetic shower, in terms of competing emission mechanisms. This radial dependence was first predicted by de Vries et al. [45] and Alvarez-Muniz et al. [114], and first measured at LOFAR [115]. As mentioned in Section 2.2.1, the radially-polarized contribution mainly originates from the charge-excess emission mechanism, so in what follows the expression *charge-excess fraction* will be used to indicate the relative contribution of radially-polarized emission with respect to the dominant geomagnetic one.

6.1 Event Selection and Reconstruction

For the following analysis, only radio events externally triggered by the AERAlet array (see Section 3.4) have been included. The data-taking period runs from the 20th of February 2013, the first stable day for AERAlet,

till the end of December 2013. However, due to the deployment of AERA phase II in March and April 2013, few events have been recorded in those months, and no events in May 2013 due to communication problems in the arrangement phase of the new radio stations.

The event reconstruction has been performed with the Auger $\overline{\text{Off}}\underline{\text{line}}$ Framework, adopting the pipeline specified in Section 4.3. As mentioned in this Section, first the shower geometry is reconstructed using the SD data, and then the information about the electric field polarization is extracted from the recorded radio traces. This information is given per each antenna flagged as a *candidate* station, i. e. as a station that likely has recorded a genuine radio pulse according to the criteria outlined in Section 4.3, and it can be used to estimate the contribution of the radially-polarized component to the total emission.

6.1.1 Antenna-wise charge-excess fraction

As reported in Section 4.3.1, in the reference frame illustrated in Fig. 4.4, the inferred polarization information is summarized for each antenna in terms of the angle of polarization ψ between the measured electric field and the $\hat{\mathbf{v}} \times \mathbf{B}_{\oplus}$ direction, chosen as the x-axis, and the observer angle ϕ between this direction and the projected antenna position. In this reference frame no component of the electromagnetic field is present along the $\hat{z}$ axis, as this axis is chosen to coincide with the propagation vector of the emitted radio wave (see Section 4.3.1 for details). For this reason, one can write the expected electric field at a given time t as a sum of the two main contributions (cf. discussion before Eq. (2.3))

$$\mathbf{E}(t) = \mathbf{E}_{\text{G}}(t) + \mathbf{E}_{\text{C}}(t) = |\mathbf{E}_{\text{G}}(t)|\hat{\boldsymbol{x}} + |\mathbf{E}_{\text{C}}(t)| \cos\phi\, \hat{\boldsymbol{x}} + |\mathbf{E}_{\text{C}}(t)| \sin\phi\, \hat{\boldsymbol{y}}, \tag{6.1}$$

where the geomagnetic component $\mathbf{E}_{\text{G}}$ is completely polarized along the $\hat{\boldsymbol{x}}$ axis. This contribution scales with the opening angle α between the propagation direction of the shower and the geomagnetic field (cf. Eq. 2.2). Taking into account this dependency, the charge-excess fraction a at the pulse-maximum time (so that one can drop the time dependency) is usually defined as [116], [54]

$$a \equiv \sin\alpha \frac{|\mathbf{E}_{\text{C}}|}{|\mathbf{E}_{\text{G}}|}. \tag{6.2}$$

Now, Eq. (6.1) above reads (again at the time for which the the field reaches its maximum amplitude and assuming that this maximum in $\mathbf{E}_{\text{G}}$

and $\mathbf{E}_C$ is reached simultaneously)

$$\begin{aligned} |\mathbf{E}| &= |\mathbf{E}_G|\hat{\boldsymbol{x}} + |\mathbf{E}_G|\frac{a}{\sin\alpha}\cos\phi\,\hat{\boldsymbol{x}} + |\mathbf{E}_G|\frac{a}{\sin\alpha}\sin\phi\,\hat{\boldsymbol{y}} \\ &= |\mathbf{E}_G|\left[\left(1 + \frac{a}{\sin\alpha}\cos\phi\right)\hat{\boldsymbol{x}} + \frac{a}{\sin\alpha}\sin\phi\,\hat{\boldsymbol{y}}\right]. \end{aligned} \tag{6.3}$$

Therefore, the ratio of these two components equals the tangent of the (polarization) angle ψ between the electric field vector and the $\hat{\mathbf{v}} \times \mathbf{B}_\oplus$ direction

$$\tan\psi = \frac{\sin\phi}{\dfrac{\sin\alpha}{a} + \cos\phi}, \tag{6.4}$$

As anticipated in Section 4.3.1, this angle as well as the observer angle ϕ are calculated in the `RdPolarizationReconstructor` module, respectively by Eqs. (4.13) and (4.15), for the candidate antennas. Then, for the i-th antenna of the j-th event, it is possible to invert the above equation to get a per-antenna value a_i for the charge-excess fraction, provided that $\sin\phi_i - \cos\phi_i \tan\psi_i \neq 0$ (which would happen in the extremely unlikely case that $\psi_i \equiv \phi_i$):

$$a_i = \frac{\sin\alpha_j \tan\psi_i}{\sin\phi_i - \cos\phi_i \tan\psi_i}. \tag{6.5}$$

As foreseen in the definition in Eq. (6.2), this expression clearly shows how the angle α_j between the shower axis and the geomagnetic field tunes the strength of the radial-polarization component through the $\sin\alpha_j$ factor.

Fig. 6.1 shows the function in Eq. (6.5) for several values of the polarization angle against the observer angle. In this plot the $\sin\alpha$ factor has been set to 1 (i. e. $\alpha = \pi/2$), as it would only change the scale. If the polarization angle of the measured electric field is close to 0 (or π), the two main emission contributions interfere destructively (or constructively). This happens when the directions of the electric fields generated by the Lorentz force and by the Askaryan mechanisms coincide, i. e. when the observer angle is $0, \pi$, or 2π. In these cases the geometry is degenerate, and as a consequence the function $a(\psi, \phi; \sin\alpha)$, shown in Eq. (6.5), diverges.

This degeneracy is reflected as well in the "sign" of the computed fraction $a(\psi, \phi; \sin\alpha)$: a shift of quadrant for the observer angle has the effect of *flipping* its sign (cf. Fig. 6.1). For clarity reasons, these considerations

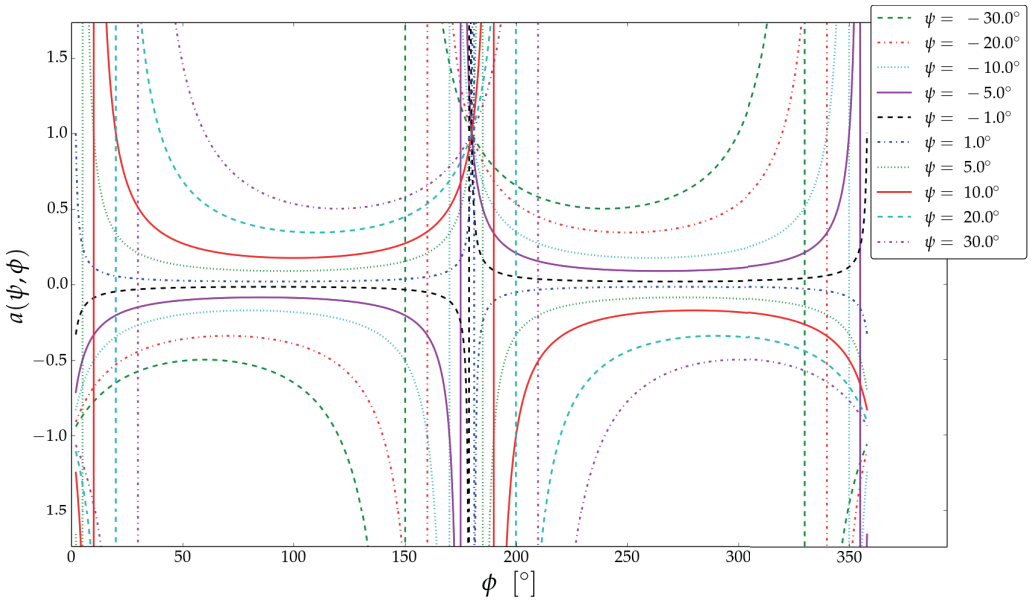

FIGURE 6.1: The charge-excess fraction a as a function of the polarization and observer angles ψ and ϕ respectively.

are schematically summarized in Table 6.1, which also highlights how a, ψ, and ϕ are interconnected in this respect. In cases where the observer angle and the polarization angle have a high chance of being ill determined, one can expect Eq. (6.5) to return 'non-physical' results.

	$\phi \in (0, \pi)$	$\phi \in (\pi, 2\pi)$
$a \gtrsim 0$	$\psi \gtrsim 0$	$\psi \lesssim 0$
$a \lesssim 0$	$\psi \lesssim 0$	$\psi \gtrsim 0$

TABLE 6.1: Conjunction of the possible values of charge-excess fraction, polarization angle, and observer angle.

The reasons above make clear that, at the station level, a selection strategy is of vital importance to minimize the presence in the sample of ill-determined angles. At the same time, to avoid selection biases, no 'straight' cut can be applied, as such a cut would unavoidably include a certain degree of arbitrariness for the choice of, for instance, the "good" cut value. For this reason, we chose to adopt another way to reject sta-

tions that likely have an ill-determined angle, either the observer angle or the polarization angle.

As it will be illustrated in Section 6.1.3, the adopted selection makes use of the angular uncertainties as well. However, the `RdPolarizationReconstructor` module does not provide the uncertainty on the observer angle, hence before going into the details of the station-selection criteria, one needs to determine this uncertainty. For this purpose, in the next Section, the observer-angle uncertainty is determined via a Monte Carlo method.

6.1.2 Observer-Angle Uncertainty

The shower-core position uncertainty represents the dominant contribution to the uncertainty of several other shower parameters. Indeed, the location of the shower core is in turn used to determine other quantities, typically the station distances on the shower plane, so any other quantity that depends on this distance will be affected as well.

On the $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ (shower) plane specified above, $\mathbf{r}_i$ is the vector from the shower core to the position of the i-th station, projected onto the shower plane, which in polar coordinates is identified by $\mathbf{r} \equiv (r, \phi)$. The observer angle ϕ is counted counterclockwise from the $\hat{\mathbf{x}}$ axis ($\equiv \hat{\mathbf{v}} \times \mathbf{B}_{\oplus}$), and accordingly determined as given in Eq. (4.15). Its uncertainty σ_ϕ, has been estimated by performing a MC simulation for the observer-angle distribution resulting from the variation of the shower-core location within its uncertainty ellipse (as estimated by the SD reconstruction). This effect is shown schematically in Fig. 6.2. As an example, two shower-core locations are taken at the extremes of the core-uncertainty ellipse. To these core positions correspond, in the shower plane, the varied antenna positions indicated in this Figure by $(r,' \phi')$, (r'', ϕ'').

For each event, and for each antenna in the event the main steps of the simulation are listed below:

- ▷ The bivariate normal p.d.f. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the core position is generated, taking $\boldsymbol{\mu} \equiv (x_C, y_C)$, and covariance matrix $\boldsymbol{\Sigma}$. Here, the estimated core position, the uncertainties and correlation between coordinates are all taken from the SD reconstruction.
- ▷ For every sample $(x_C^{(s)}, y_C^{(s)})$ of the generated p.d.f., the vector $\mathbf{r}^{(s)}$ is derived and the corresponding observer angle $\phi^{(s)}$ is calculated by means of Eq. (4.15).

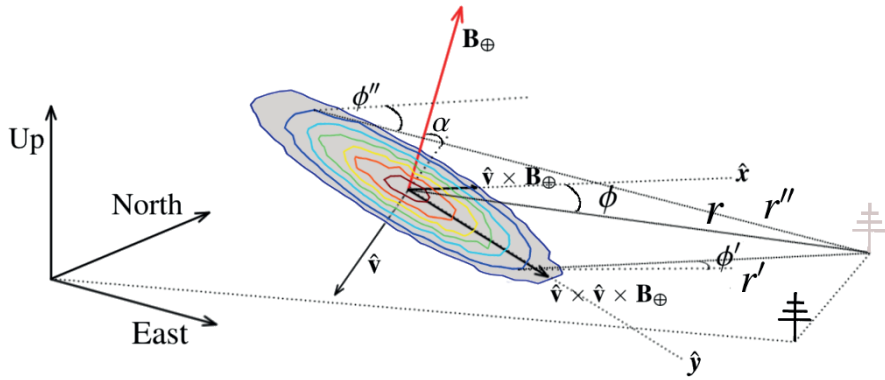

FIGURE 6.2: Schematic illustration of the influence of the core-location uncertainty on the determination of the observer angle. In the used reference frame an antenna position is projected onto the shower plane (in grey); in polar coordinates identified by (r, ϕ). The variation of the core-location within its uncertainty ellipse results in the variation of the antenna polar coordinates. This effect is shown, in an amplified way, by the varied coordinates $(r,' \phi')$ (r'', ϕ''). Artist impression of the shower plane with the antenna lifted up the shower plane (in gray) to better visualize this effect.

▻ The variance σ_ϕ^2 can then be estimated from the resulting distribution, making use of the properties of the Von Mises distribution, which approximates well enough the wrapped normal distribution on S^1 (cf. Appendix B).

Fig. 6.3 shows the estimated observer-angle uncertainty as a function of the planar distance from the shower core, as a binned profile both in uncertainty-ellipse 'extension' ($\pi\sigma_X\sigma_Y$) and distance. It is evident that the influence of the core uncertainty on the determination of the observer angle fades down with the distance of the station from the shower core. At the same time, for close stations, the extension of the core-uncertainty ellipse determines the spread of the resulting distribution of calculated observer angles. The plot also shows an empirical five-parameter model fitted to the MC-generated data, based on the above considerations:

$$\sigma[\phi_i] = \left(a \cdot \pi\sigma_{X_i}\sigma_{Y_i} + \frac{b}{r_i + c} \right) \exp\{-d \cdot r_i\} + e \ \ [^\circ]. \tag{6.6}$$

To fit this model, an additional step is needed to determine the uncertainty $\sigma[\sigma_\phi]$, as a weight for the target function. For this purpose, the procedure to calculate σ_ϕ is repeated hundred times to generate a distri-

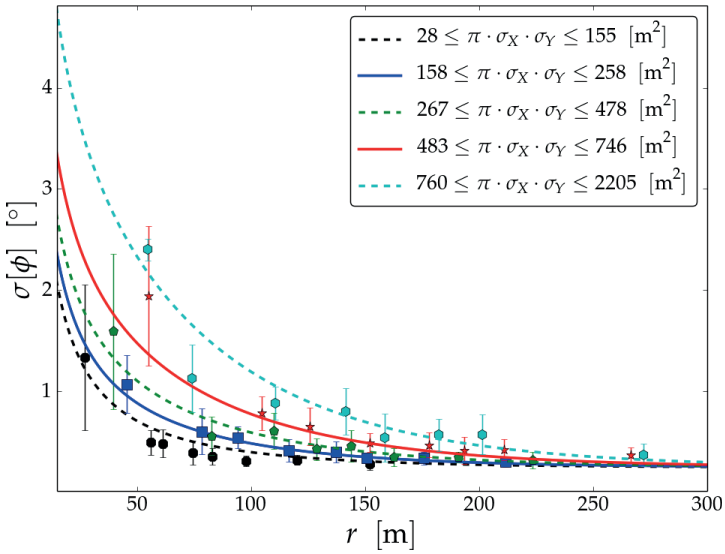

FIGURE 6.3: Binned profile of the estimated observer-angle uncertainty as a function of station distance from the shower-core. The different colours indicate the 'extension' of the core-uncertainty ellipse as in the legend.

bution from which the standard deviation $\sigma[\sigma_\phi]$ can be estimated. The least-squares minimum is reached for the set of parameters listed in Table 6.2.

Parameter	a	b	c	d	e
Value	0.0032	33	2.2	0.014	0.24
Error	0.0004	3	0.7	0.001	0.02

TABLE 6.2: Parameters for the observer-angle uncertainty model specified in Eq. (6.6), determined by a least-squares fit.

The model given by the parameters in Table 6.2 is then used in the following, to provide the uncertainty for the observer angle.

As already mentioned, one needs to apply the quality cuts at the event level to insure that all radio-pulses detected are associated with a physical cosmic-ray event. At the station level, quality cuts are intended to minimize the presence of transient-noise pulses and to avoid unstable regions where the determination of the charge-excess fraction a is problematic.

6.1.3 Event and Station Selection

The event-wise quality cuts have been applied on the SD-part of the combined reconstruction:

- ◦ Zenith-angle cut: $\theta_{\mathrm{SD}} \leqslant 50^{\circ}$.
- ◦ Events that fulfil the posterior T5 condition.
- ◦ Energy above the AERAlet energy threshold $E_{\mathrm{SD}} > 0.04$ EeV.

After these cuts, 251 events are left out of 1025 totally reconstructed ones, with a total of 2971 radio signals not rejected, $S/N > 10$ (cf. Section 4.3). After that, for consistency reasons, we only include in the analysis measurements performed with AERA phase I RDS's, which reduces the number of selected radio signals to 855. Moreover, as we have selected T5 events using the AERAlet array, the core is fully contained in this hexagon and therefore the AERA I stations are the closest ones to the shower core.

The following quality cuts have been applied to these remaining 855 radio signals in the dataset after the event selection above:

- ⋆ Maximum amplitude found in the noise region $< 200\,\mu\mathrm{V}\cdot\mathrm{m}^{-1}$. Usually, transient-noise contaminated traces present multiple pulses of this sort, being the noise-level amplitude of the order of few tens of $\mu\mathrm{V}\cdot\mathrm{m}^{-1}$. A high amplitude noise level is a clear feature of such a contaminated trace.
- ⋆ Finite observer-angle uncertainty, based on the method described in Section 6.1.2. This method might not converge if the station is contained in the shower-core uncertainty ellipse.
- ⋆ Reliable observer angle determination (see below).
- ⋆ Reliable polarization angle determination (see below).

As briefly anticipated above, the use of Eq. (6.5) for the determination of the charge-excess fraction, suggests to avoid ill-determined angular variables that would result in unreliable values for a. In particular, let us first consider the observer angle ϕ, and let ϕ' and $\sigma^2_{\phi'}$ be respectively its estimated value and its variance. To determine whether the true value ϕ^* has a significant probability to be in the next quadrant with respect to the estimated one, we test whether the distribution of its cosine can be approximated with a normal distribution. In fact, as $\phi' \longrightarrow 0, \pi, 2\pi$ the

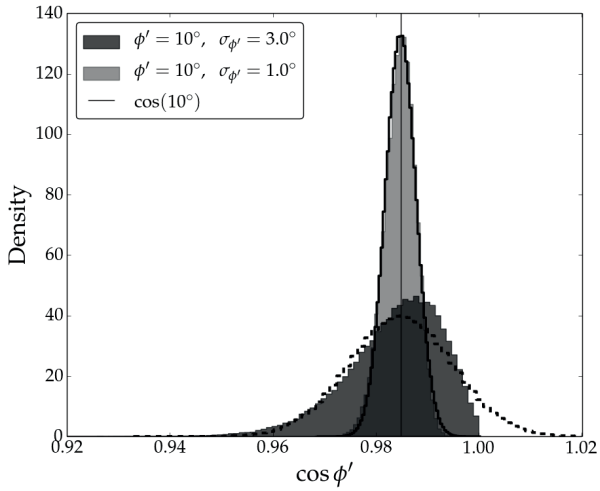

FIGURE 6.4: Examples of generated distributions for the cosines of an observer angle $\phi' = 1°$ but two different values of the uncertainty – cf. Eq. (6.7) – as indicated in the legend. To guide the eye, the vertical line is drawn precisely at $\cos(10°)$. The black-solid and black-dashed lines represent two genuine normal distributions to highlight the departure from normality.

transformed distribution

$$\cos(\phi'_1, \ldots, \phi'_M | \phi', \sigma_{\phi'}), \quad \text{where} \quad \phi'_1, \ldots, \phi'_M | \phi', \sigma_{\phi'} \sim \mathcal{N}(\phi', \sigma_{\phi'}). \tag{6.7}$$

becomes asymmetric. Though, if the uncertainty $\sigma_{\phi'}$ is "sufficiently small", the distribution of cosines will stay symmetric even if ϕ' is, for instance, very close to 0. As an example, Fig. 6.4 displays the distribution in Eq. (6.4) generated for $\phi' = 10°$, but two different values of $\sigma_{\phi'}$), respectively $1°$ (light gray) and $3°$ (dark gray), further compared to two genuine normal samples (solid-black lines). As expected, the asymmetry is more evident for the generated $\cos\phi'$ distribution with the largest uncertainty.

To test if the transformed distribution is compatible with a normal one we have used the Anderson-Darling test [117], which has been found to be one of the best empirical-distribution-function statistics for detecting most departures from normality [118].

Summarizing, for the i-th antenna participating in the j-th event, a normal distribution is generated $\mathcal{N}(\phi'_i, \sigma[\phi'_i])$ of $M_\phi = 10000$ samples. Then, the corresponding distribution of cosines is considered and tested through the Anderson-Darling test. The radio signal is rejected if the test statistic A^2 is larger than all the critical values returned by the test if normality

holds, at the 15%, 10%, 5%, 2.5%, and 1% significance level, selected if the test statistic is lower than at least one critical value. In other words, the station is rejected if one cannot find a 'reasonable' significance level for which the corresponding critical value is smaller than the test statistics. Fig. 6.5 shows the effect of this cut criterion on the observer angle distribution. It can be seen that, we have avoided the choice of a "forbidden" range around $0, \pi, 2\pi$, and also the arbitrariness in the choice of a "small enough" uncertainty.

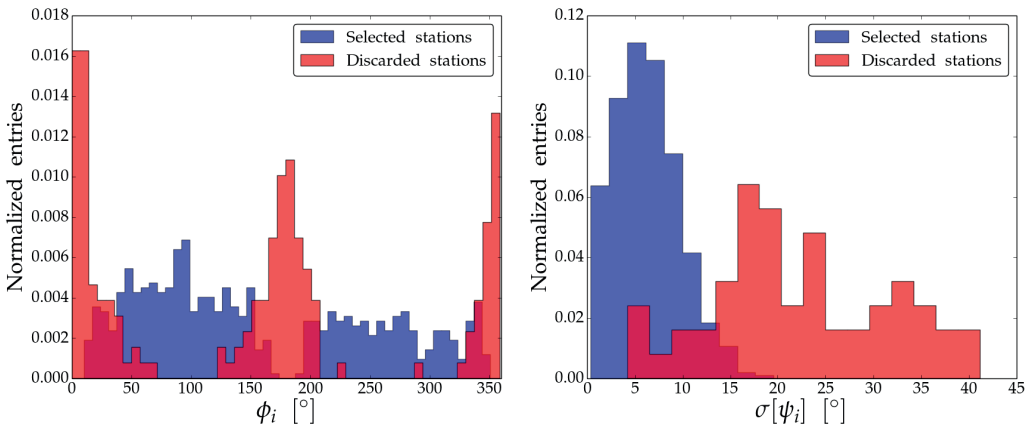

FIGURE 6.5: *Left panel:* distribution of the observer angle ϕ for selected (blue histogram) and rejected (red histogram) radio signals after the normality-check cut. The visible high density for the set of rejected signals, is caused by having normalized the two data sets separately. *Right panel:* same as left but for the polarization-angle uncertainty σ_ψ.

Concerning the polarization angle ψ, given ψ' and $\sigma^2_{\psi'}$ a certain estimate of its variance, the test has been repeated for its tangent

$$\tan(\psi'_1, \ldots, \psi'_{M_\psi} | \psi', \sigma_{\psi'}), \quad \text{where} \quad \psi'_1, \ldots, \psi'_{M_\psi} | \psi', \sigma_{\psi'} \sim \mathcal{N}(\psi', \sigma_{\psi'}). \tag{6.8}$$

For small angles the tangent is approximated by the angle and there is no departure from normality. As ψ grows, the non-linearity of $\sin\psi$ starts to have its weight on the tangent, but at the same time this is counterbalanced by $\cos\psi$, therefore only angles with an outstanding uncertainty will show a distribution departing from a normal one. Fig. 6.5-*right* shows the effect of this cut criterion on the polarization-angle distribution. It is interesting to notice that some stations with large polarization-angle uncertainty have quite a high S/N pulses, which is something unexpected considering how the polarization information is extracted from

the trace. Most likely, these pulses have been recorded with unstable atmospheric conditions, such as thunderstorms, that in absence of weather-monitoring data were still present in the data set. In such a case, the presence of strong atmospheric electric fields will affect the dominant polarisation direction of the CR-induced radio signal[1].

Fig. 6.6 shows the polarization angle as a function of the observer angle both for selected and discarded stations. As expected from Eq. (6.4), a sinusoidal-like pattern is clearly visible, as the angle of polarization depends on the observer position, which is a clear indication that the charge-excess mechanism does contribute to the measured total electric field. After these additional cuts, 539 radio signals are considered for the following analysis.

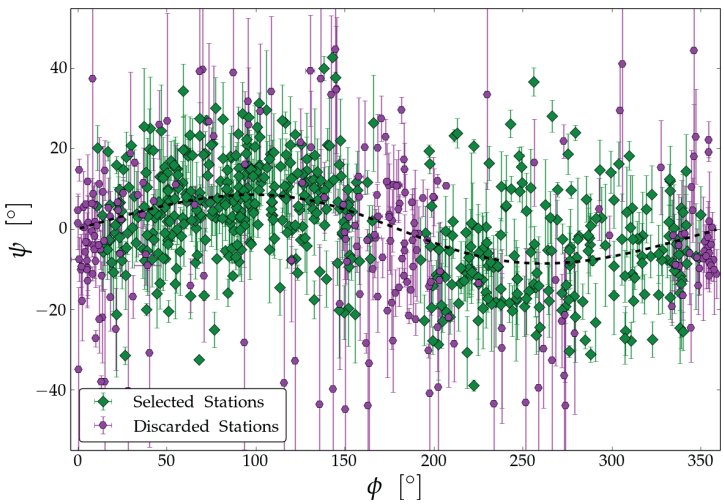

FIGURE 6.6: The angle of polarization ψ as a function of the observer angle ϕ shows the expected sinusoidal-like pattern. The black-dashed line is drawn to guide the eye.

6.2 Modelling the amount of Radial Polarization

As mentioned above, it has been recently shown by LOFAR that the fraction of radially-polarized emission relatively to the dominant geomagnetic one increases with the radial distance from the shower axis at least up to 250 m.

At Auger, the use of AERAlet as an external trigger for the radio array,

[1]Recently, this effect has been also observed [119].

allows a further investigation of the lateral dependence of the contribution of the radially-polarized component in the region close to the shower core. Indeed, the shower-core location of the events passing the quality cuts is fully contained in the AERAlet cell, which not accidentally overlaps with the phase-I stations of AERA. As a consequence, these antennas will record a signal coming from the central region of the wave front.

Following a multi-event approach, i. e. collecting together the stations from all events, one expects this region to be sampled in quite a regular way. Of course here, one is making the tacit assumption that the radio wavefront, at least close to the shower core, can be considered *universal* in terms of competing radio-emission mechanisms, i. e. in first approximation the fall-off of the charge-excess fraction with increasing radial distance has a *similar* behaviour independently of the shower geometry. In other words, close to the core (up to 250 m), changes of the charge-excess fraction due to the different signal fall-off of the two main emission mechanisms are of lower order than changes induced by other event/shower parameters. Ultimately, we want to check to which extent the radio facility at Auger has enough sensitivity to observe the predicted initial rises of the charge-excess fraction a with increasing radial distance close to the shower axis.

For this purpose, first the uncertainty on ψ and ϕ need to be propagated to assign the uncertainty to the corresponding values $a_i = a(\psi_i, \phi_i | \sin \alpha_j)$. The uncertainty on a has then been calculated by a MC method, assuming ψ and ϕ to be independent, yielding the distributions

$$a_i^{(1)}, \ldots, a_i^{(S)} | \psi_i, \sigma_{\psi_i}, \qquad \text{and} \qquad a_i^{(1)}, \ldots, a_i^{(S)} | \phi_i, \sigma_{\phi_i}, \tag{6.9}$$

for each station i. These distributions can be approximated by drawing a number $S = 10000$ of samples from $\mathcal{N}(\psi, \sigma_\psi)$, and $\mathcal{N}(\phi, \sigma_\phi)$, and for each of these samples calculating a by making use of Eq. (6.5). From each of the generated distributions in Eq. 6.9, one can extract the sample variances $\sigma^2_{a_i|\psi_i}$, $\sigma^2_{a_i|\phi_i}$, which respectively indicate the uncertainty in a due to the uncertainty in ψ, and ϕ. Finally, the sum $\sigma^2_{a_i|\psi_i} + \sigma^2_{a_i|\phi_i}$ of these variances has been assigned as the variance $\sigma^2_{a_i}$ of a_i for the i-th selected radio signal. Due to the additional cuts in Section 6.1, now the uncertainty on ψ dominates as we observe that $\sigma^2_{a_i|\psi_i} \gtrsim 10 \times \sigma^2_{a_i|\phi_i}$, which implies that $\sigma^2_{a_i|\phi_i}$ does almost not contribute to the final propagated uncertainty on a ($\lesssim 1\%$).

For eight signals in the data set, the generated distribution $\mathcal{N}(\psi, \sigma_\psi)$ is

quite broad, and due to the use of Eq. (6.5), the obtained value of σ_a is extremely large (> 10). However, these signals have been kept in the following analysis.

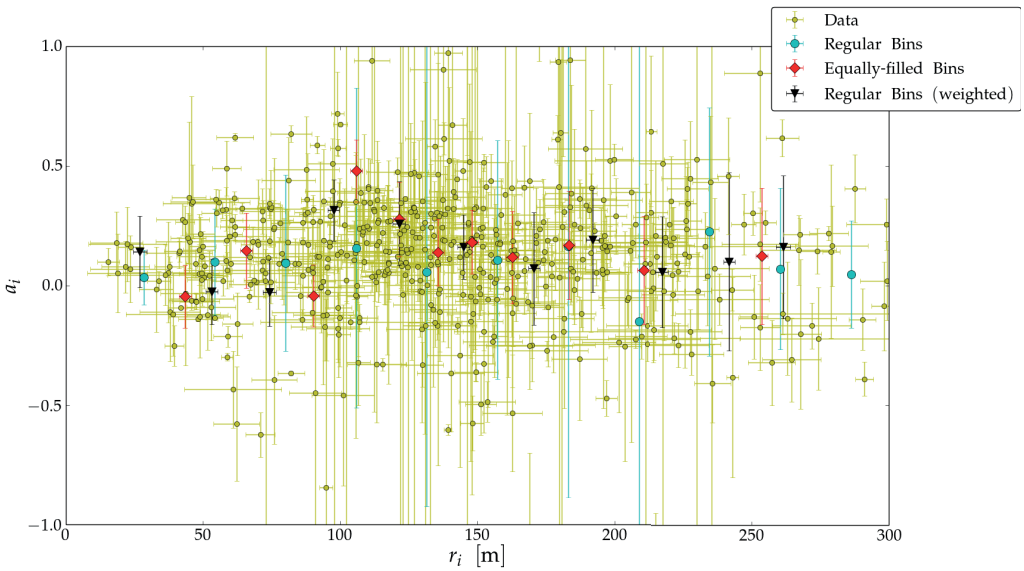

FIGURE 6.7: The station-wise charge-excess fraction a_i (yellow points), as given by Eq. 6.2, as a function of the station distance from the shower axis. Together with the single data points, three profiles are also shown, for three different ways of binning the data (see legend and text for details).

Fig. 6.7 shows the calculated value a_i of the charge-excess fraction, calculated from Eq. (6.5), versus the shower-plane distance of the antenna from the shower axis, for each selected station. Together with the single data points, the plot shows the average in eleven regularly-spaced bins (blue circles), a weighted average (defined in Eq. (6.10)) in eleven equally-filled bins (red diamonds), and a weighted average in eleven regularly-spaced bins (black triangles).

$$\bar{a}_k = \frac{1}{\sum_{l=1}^{N_k} \frac{1}{\sigma^2[a_l]}} \sum_{l=1}^{N_k} \frac{a_l}{\sigma^2[a_l]}, \qquad \sigma[\bar{a}_k] = \frac{1}{\sum_{l=1}^{N_k} \frac{1}{\sigma^2[a_l]}} \sum_{l=1}^{N_k} \frac{\sigma[a_l]}{\sigma^2[a_l]}. \quad (6.10)$$

The eight outliers with $\sigma_a > 10$ are not visible as for these radio signals $|a| > 1$. Due to their large uncertainties, the weight of the outliers in the averaging process is negligible.

As a constant a-fraction was found in [54], a constant $\mathfrak{a}$ has been fitted

through the data, returning $\mathfrak{a} = 0.135 \pm 0.002$ (with $\chi^2_{\text{red}} = 42.9$). At the same time, due to the large reduced χ^2, one can attempt to fit a distance-dependent model. A rather general distance-dependent model can be written as

$$\hat{a}(r) = \alpha_0 + \alpha_1 \cdot r^{\,\alpha_2}, \tag{6.11}$$

where α_0 is a charge-excess fraction constant contribution, and α_1, α_2 controlling the growth of a with the shower-axis distance r (cf. Table 6.3).

Parameter	α_0	α_1	α_2	χ^2_{red}
Value	-0.50	0.19	0.26	40.6
Error	0.19	0.13	0.08	

TABLE 6.3: The fitted parameters for the model given in Eq. (6.11).

Fig. 6.8 shows the results for the fitted model of Eq. (6.11) (solid-green line) and for the fitted constant $\mathfrak{a}$ (dashed-blue line), against the weighted-binned profiles introduced above (in the same figure are shown as well the models in Equations (6.12), (6.13), and (6.14) introduced below). Due to the wide scattering of the single data points, the resulting reduced χ^2, both for the model and the constant, are quite large. Though the reduced χ^2 are very similar, this scattering seems to suggest that the assumption of neglecting zenith-angle dependencies does not hold.

Indeed, as found in [115], the charge-excess fraction decreases with increasing zenith angle, and one can attempt to add a parameter to the model in Eq. (6.11) to take into account this eventual zenith-angle dependency (cf. Fig .6.8):

$$\hat{a}(r, \theta) = \alpha_0 + (\alpha_1 + \alpha_2 \cos\theta) \cdot r^{\,\alpha_3}, \tag{6.12}$$

with parameters, determined by a χ^2–minimization fit, listed in Table 6.4.

Parameter	α_0	α_1	α_2 [m]	α_3	χ^2_{red}
Value	-8.52	7.13	0.69	0.03	39.1
Error	20.97	20.69	0.20	0.06	

TABLE 6.4: The fitted parameters for the model given in Eq. (6.12).

The result of the fit is shown as well in Fig. 6.8, for $\theta = 0^\circ$ (black-dashed line) and $\theta = 50^\circ$ (black-dot-dashed line). Although the fit of this model results in a slightly smaller χ^2_{red} than the one for the model in Eq. (6.11), the reduced χ^2 is still quite large. In addition, the uncertainties of the

α_0 and α_1 parameters are very large and correlated, which does suggest that the data are not enough to constraint the fit. For this reason, one can attempt to reduce number of parameters in this model by considering

$$\hat{a}(r,\theta) = \alpha_0 + \alpha_1 \cos\theta \cdot r^{\,\alpha_2}. \tag{6.13}$$

The minimization of the χ^2 for this model returns the parameters in Table 6.5. As it can be seen from Fig. 6.8, its behaviour is very similar to the models in Equations (6.11) and (6.12).

Parameter	α_0	α_1	α_2	χ^2_{red}
Value	−0.49	0.13	0.39	39.5
Error	0.02	0.01	0.01	

TABLE 6.5: The fitted parameters for the model given in Eq. (6.13).

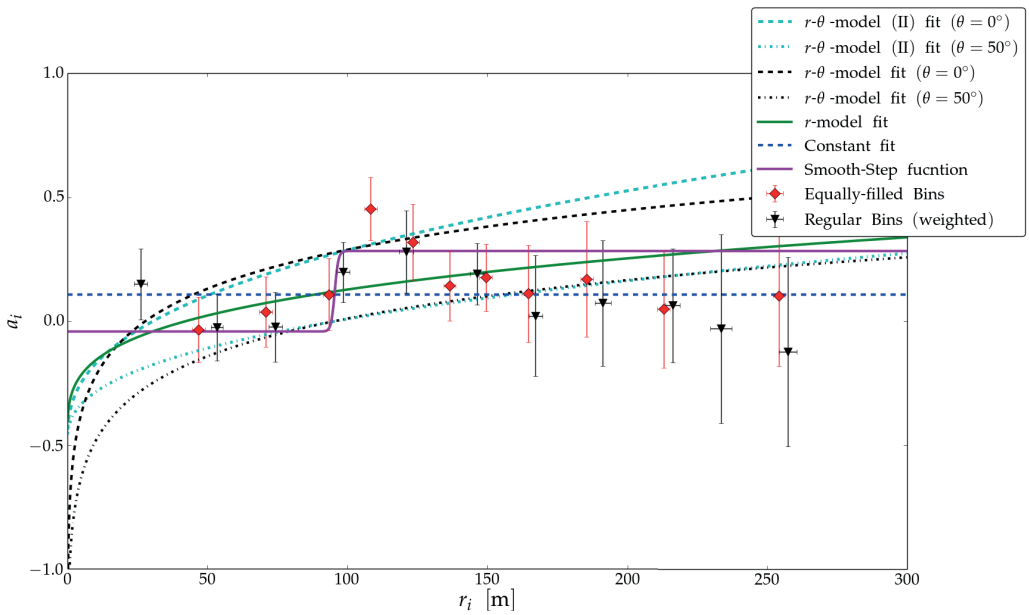

FIGURE 6.8: The fitted models (see legend) compared to the data reported in shower-axis distance bins. The r-θ models, respectively Equations (6.12) and (6.13), are shown for two fixed values of the zenith angle, namely $\theta = 0^\circ$ (dashed-black and dashed-cyan lines) and $\theta = 50^\circ$ (dot-dashed-black and dot-dashed-cyan lines), as all other curves for $\theta \in (0^\circ, 50^\circ)$ are enclosed between these two. Two different ways of binning the data have been used (see legend and text for details), though for each of those, the data point and the corresponding uncertainty are equivalently given by Equations (6.10).

Another way the data can be modelled is to consider a modified version

of the logistic function (see e. g. [120]), with four parameters [121]:

$$\hat{a}(r) = \frac{\alpha_0 \cdot \mathrm{e}^{\alpha_2 \alpha_3} + \alpha_1 \cdot \mathrm{e}^{\alpha_3 r}}{\mathrm{e}^{\alpha_2 \alpha_3} + \mathrm{e}^{\alpha_3 r}}. \tag{6.14}$$

This function models a smooth growth from α_0 to α_1, with slope α_3 and location of the middle-step point α_2, which again represents a sort of *critical distance*. Though in general the limit for $r \to 0$ of this function does not converge to α_0, it can be easily seen that it does as far as $\alpha_2 \alpha_3 \gg 1$ (cf. Table 6.6).

Parameter	α_0	α_1	α_2	α_3	χ^2_{red}
Value	−0.04	0.28	95.50	1.55	37.7
Error	0.003	0.003	0.11	0.24	

TABLE 6.6: The fitted parameters for the model given in Eq. (6.14).

Again, the resulting χ^2_{red} is smaller than the one resulting from the fit of the *r-θ* model in Eq. (6.12), though the wide spread of the data does not allow a clear judgment.

Although no definitive conclusion can be drawn from the comparison of the resulting χ^2_{red}, for each of the models introduced above one can calculate the residuals

$$\mathrm{Res}^{\mathrm{K}}(a_i) := \frac{a_i - \hat{a}_i^{\mathrm{K}}}{\sigma_{a_i}} \qquad i = 1, \ldots, N_{\mathrm{stations}}, \tag{6.15}$$

where $\hat{a}_i^{\mathrm{K}}$ indicates the model expectation for $\mathrm{K} \in \{constant, r\text{-}only, r\text{-}\theta, \text{Smooth-Step}\}$. Fig. 6.9 shows the residuals collected in eleven shower-axis-distance regular bins, for every K as indicated in the legend. For each bin, the data point represents the average of the residuals in that bin, and the error bar the 1σ spread. In proximity of the shower core, a varying charge-excess fraction with the radial distance is more consistent with the data, though this region seems to be dominated by uncertainties as highlighted by the wide spread in these bins. Here, the inclusion of the zenith angle in the *r-only* model does not improve the average residuals, though reduces its spread. On the other hand, the constant charge-excess fraction provided by the fitted constant $\mathfrak{a} \simeq 0.13$ or the fitted $\alpha_0 \simeq 0.01$ of the smooth-step function, is either too large (negative residuals) or too small (positive residuals).

Around 100 m, the expected charge-excess fraction is in agreement with the data, though large fluctuations are present and the spread in this bin

stays basically constant independently from the model. Beyond 100 m, a constant charge-excess fraction seems to be favoured over all the other the models.

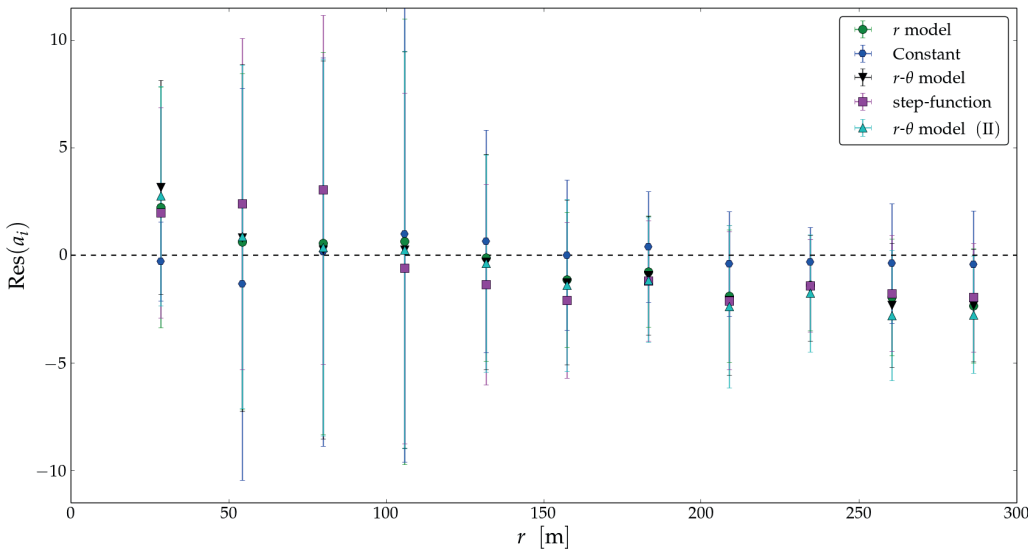

FIGURE 6.9: The mean of the residuals of the expected charge-excess fraction as in Eq. (6.15) per distance bin i, for the different models introduced in the text (see legend).

However, due to the large reduced χ^2 values, the statistical significance of such a conclusion is not sufficient to prefer one of the proposed models, especially close to the core where uncertainties dominate. To check if our assumption is reasonable, i. e. the core of AERA is sensitive enough to appreciate the variation of the charge-excess fraction with the observer distance, one can use a bootstrapping method to estimate the uncertainty on the model parameters in a more stable way, as the single χ^2 fitting does not seem to return a reliable values for the uncertainties. The details of how such an approach has been used to test our hypothesis are illustrated in the next Section.

6.2.1 Bootstrap Approach to estimate the model-parameter uncertainties

In general, as mentioned in Section 5.3.1, in bootstrapping methods the data sample is treated as the population, and by *resampling with replacement* the data many times, one can approximate the sampling distribution of a chosen statistic. In other words, each time a new sample is generated from the original one, by randomly drawing with replacement the

same number of elements, this sample will likely contain some original data points more than once and will not include some others. As a result, a certain statistic calculated from one of the new samples will have a slightly different value from the same statistic calculated from the original sample. In this way, the distribution of the statistics calculated from the resampled data sets is an estimate of the sampling distribution of the chosen statistic [101], [122].

In our case, a bootstrapping method can be used to construct the distributions for the different model parameters α_j^{K} ($\mathrm{K} \in \{constant, r\text{-}only, r\text{-}\theta, \text{Smooth-Step}\}$) and for the constant assumption $\mathfrak{a}$, from which the weighted average and uncertainty can be extracted. Hence, these estimates are taken as the final estimates to compare the models with the data and against the constant assumption.

As for the i-th selected station one needs the associated measured quantities

$$(a_i, \sigma_{a_i}, r_i, \theta_i) \qquad i = 1, 2, \ldots, N_{\text{stations}}, \tag{6.16}$$

one can resample from the set of selected stations to maintain this correspondence in the bootstrapped distributions. Hence, the bootstrapping procedure consisted of the following steps:

(*1*) N_{stations} integers s are drawn from the discrete uniform distribution $\mathcal{U}\{1, N_{\text{stations}}\}$, so that $\mathcal{S} = \{s\}$ is the new set of stations with repetition and same number of elements as the original selected set.

(*2*) From the empirical distribution of the charge-excess fractions $\{a_i\}_{i=1}^{N_{\text{stations}}}$, the resampling $\{a_s^*\}_{s\in\mathcal{S}}$ is generated with associated uncertainties, distances, and zenith angles $\{\sigma_{a_s}^*, r_s^*, \theta_s^*\}_{s\in\mathcal{S}}$, from the respective empirical distributions.

(*3*) The models in Equations (6.11), (6.12), (6.14), and a constant are fitted to the resampled data, to obtain a realization of the parameters α_j^* and of the constant $\mathfrak{a}^*$, with corresponding uncertainties.

(*4*) These steps are then repeated $M = 10^4$ times, to construct the approximating distributions $\{\alpha_{j,k}^*\}_{k=1}^M$, $\{\mathfrak{a}_k^*\}_{k=1}^M$ for the model parameters, and for the constant.

(*5*) These distributions, $\{\alpha_{j,k}^*\}_{k=1}^M$, and $\{\mathfrak{a}_k^*\}_{k=1}^M$, are then used to estimate the weighted average and uncertainties to be compared with the single-fit results obtained in the previous Section.

Such a procedure results in the distribution shown in Fig. 6.10 when

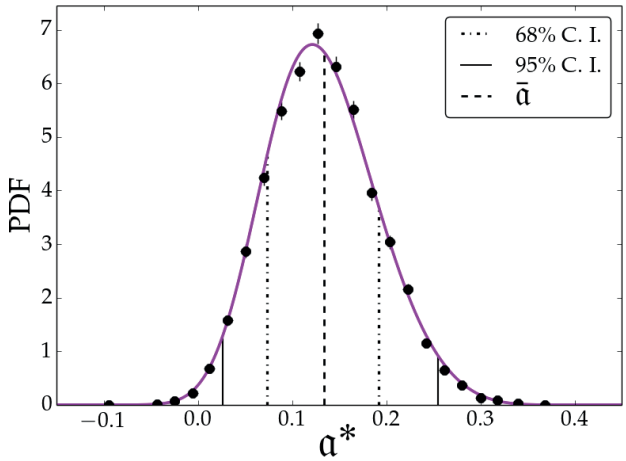

FIGURE 6.10: The approximating distribution $\{a_k^*\}_{k=1}^M$, from which the confidence intervals are extracted after fitting a suitable p.d.f. (solid-magenta line). The weighted average ā (dashed line) coincides with the mean of the fitted density.

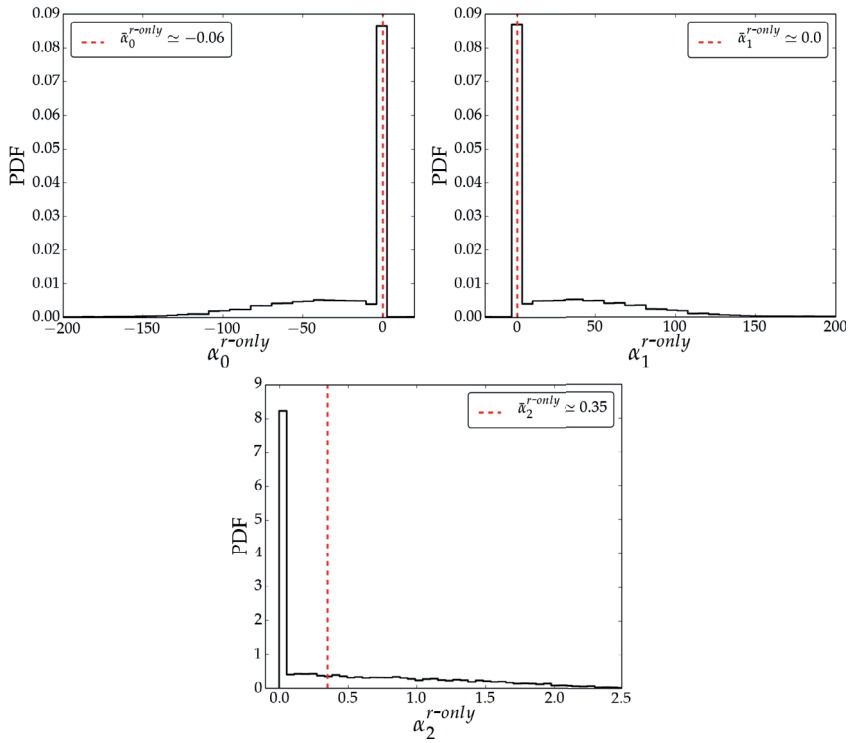

FIGURE 6.11: The $\{\alpha_{j,k}^*\}_{k=1}^M$ approximating distributions ($j = 0, 1, 2$) for the *r-only* model in Eq. (6.11), resulting from the bootstrapping method, together with their weighted average indicated by the dashed-red line.

fitting a constant through the data. This Figure shows the distribution of the bootstrapped constant fits (black dots), modelled as a generalised extreme-value continuous random variable (solid-magenta line). In this figure the dot-dashed (solid) lines indicate the 68% (95%) confidence interval $[0.07, 0.19]$ ($[0.03, 0.25]$), deduced from the fitted density. The averaged constant value $\bar{\mathfrak{a}} = 0.13$ (indicated by the dashed line) coincide with the mean of this distribution, and is consistent with the single-fit result determined in the previous Section.

Concerning the *r-only* model in Eq. (6.11), Fig. 6.11 summarizes the results of the bootstrapping procedure. For each of these distributions, the weighted average is indicated by the red-dashed line. The weighted average of the $\alpha_1^{r\text{-}only}$ parameter is essentially zero (cf. 6.7), restraining any distance-dependent effect. At the same time $\alpha_0^{r\text{-}only}$ is very small as well, as also evident from this Figure, which highlights

Parameter	$\alpha_0^{r\text{-}only}$	$\alpha_1^{r\text{-}only}$	$\alpha_2^{r\text{-}only}$ [m]
Value	-0.06	$2.8 \cdot 10^{-8}$	0.35
Error	0.09	$1.7 \cdot 10^{-4}$	0.28

TABLE 6.7: Parameter values and uncertainties (weighted averages) for the model given in Eq. (6.11) after the bootstrapping procedure.

For the *r*-θ model in Eq. (6.13), the parameter distributions are shown in Fig. 6.12, and the corresponding weighted averages and uncertainties reported in Table 6.8. Also for this model $\alpha_1^{r\text{-}\theta} \simeq 0.0$, and again no distance dependency is displayed.

Parameter	$\alpha_0^{r\text{-}\theta}$	$\alpha_1^{r\text{-}\theta}$	$\alpha_2^{r\text{-}\theta}$
Value	-0.23	$2.4 \cdot 10^{-7}$	0.23
Error	0.08	$4.1 \cdot 10^{-4}$	0.07

TABLE 6.8: Parameter values and uncertainties (weighted averages) for the model given in Eq. (6.13) after the bootstrapping procedure.

Finally, for what concerns the smooth-step model, the bootstrapping procedure returns the parameters summarized in Table 6.9. In particular, $\alpha_0^{\text{Smooth-Step}} \approx 0$ and $\alpha_1^{\text{Smooth-Step}} \simeq 0.24$, as for the single-fit results (see Table 6.6). This model provides a smooth transition from a negligible charge-excess fraction close to the shower core, where the geomagnetic contribution effectively dominates, up to an appreciable non-zero fraction farther away from the core. The middle-point of the step has been

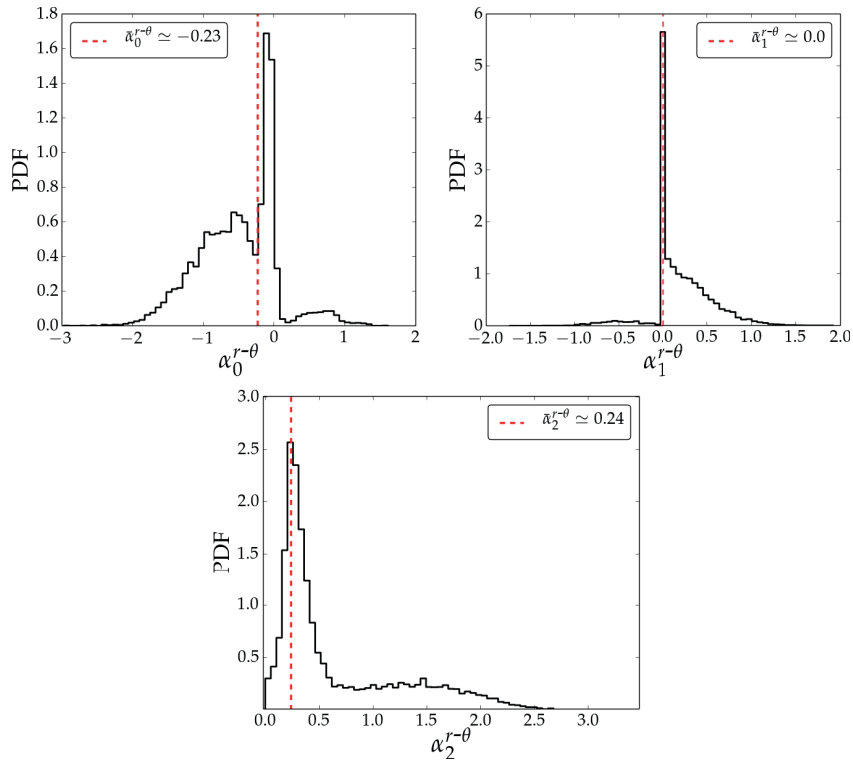

FIGURE 6.12: The $\{\alpha^*_{j,k}\}_{k=1}^{M}$ approximating distributions ($j = 0, 1, 2$) for the r-θ model in Eq. (6.13), resulting from the bootstrapping method, together with their weighted average indicated by the dashed-red line.

estimated to $\alpha_2^{\text{Smooth-Step}} \simeq 87$ m, again in agreement with the single-fit result. On the other hand, the growing rate of the charge-excess fraction $\alpha_3^{\text{Smooth-Step}} \simeq 0.05$ is much smaller than the previous estimate ($\alpha_3 \simeq 1.55$ in Table 6.6), as it is also visible comparing the two magenta curves before (Fig. 6.8) and after the bootstrap (Fig. 6.14).

Parameter	$\alpha_0^{\text{Smooth-Step}}$	$\alpha_1^{\text{Smooth-Step}}$	$\alpha_2^{\text{Smooth-Step}}$ [m]	$\alpha_3^{\text{Smooth-Step}}$
Value	-0.03	0.24	87.0	0.05
Error	0.05	0.05	0.4	0.08

TABLE 6.9: The fitted parameters for the model given in Eq. (6.14) after the bootstrapping procedure.

Fig. 6.14 shows a comparison between the binned data and the proposed models plotted using the weighted averages over the approximating distributions. In this Figure, it is evident that only a constant charge-excess

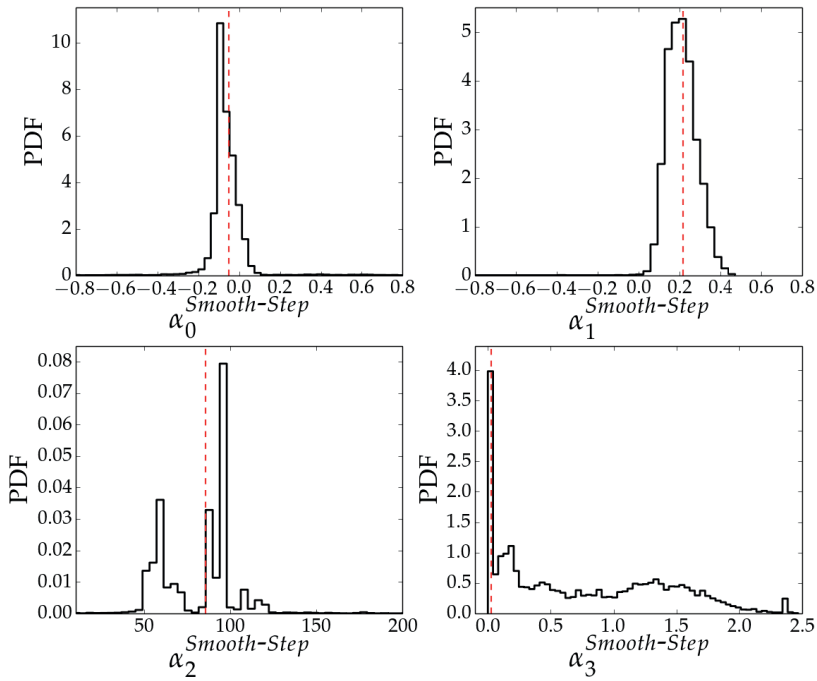

FIGURE 6.13: The $\{\alpha^*_{j,k}\}_{k=1}^{M}$ approximating distributions ($j = 0, \ldots, 3$), for the Smooth-Step model in Eq. (6.14), resulting from the bootstrapping method, together with their weighted average indicated by the dashed-red line.

fraction and a smooth growth of the charge-excess fraction with distance are consistent with the data.

As a consequence, one can use the most robust estimates of the parameters, namely the constant $\bar{a}$ and $\{\alpha^{\text{Smooth-Step}}\}_{j=0}^{4}$, to calculate the χ^2_{red}. As we find $\chi^2_{\text{smooth}-\text{step}} \simeq 36$ and $\chi^2_{\text{constant}} \simeq 39$, it is natural to repeat again the comparison of the residuals as a function of the shower-axis distances, as such large reduced χ^2 are due mainly to the wide spread of the data. Fig. 6.15 shows the calculated residuals as in Eq. (6.15), with the parameter estimated by the bootstrapping procedure. The residual distribution has been first split into fourteen regular shower-axis-distance bins, then for each bin the plotted point and error bar respectively represent the mean and the 1σ spread in that bin.

The overall picture does not differ much from the previous one (cf. Fig 6.9), the data in the region close to the core are widely spread, whose effect is reflected in the large spread of the residuals. If on one side, fluctuations prevail up to $\approx$ 70 m (between the third and the fourth bin). On the other side, the increase of the charge-excess fraction between $\approx$ 70 and 130 m,

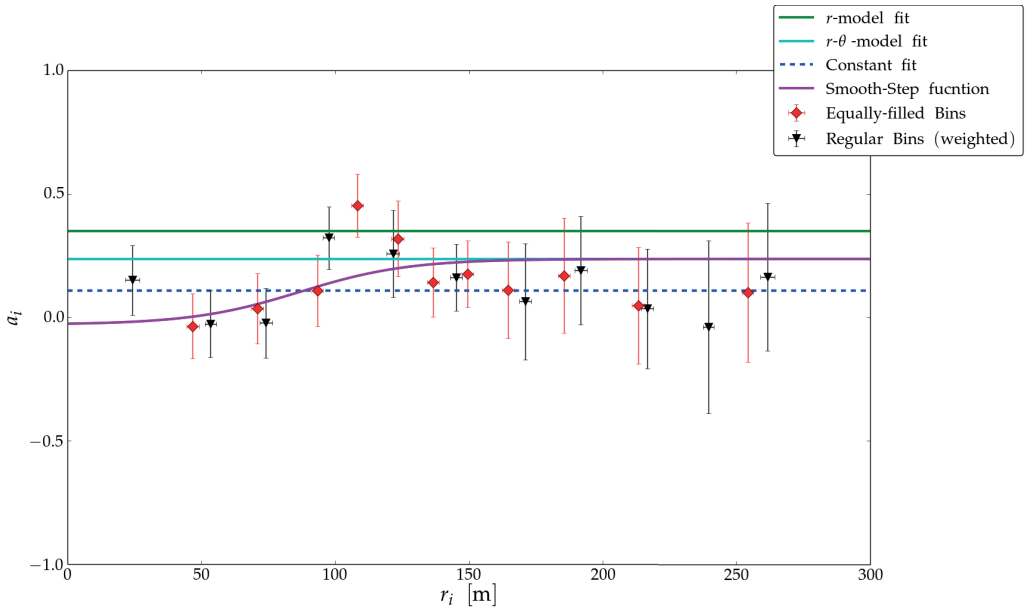

FIGURE 6.14: Comparison between the models (see legend) after bootstrapping the fit procedures, and the data, reported in shower-axis distance bins. Two different ways of binning the data have been used (see legend and text for details), though for each bin, the data point and the corresponding uncertainty are given by Equations (6.10).

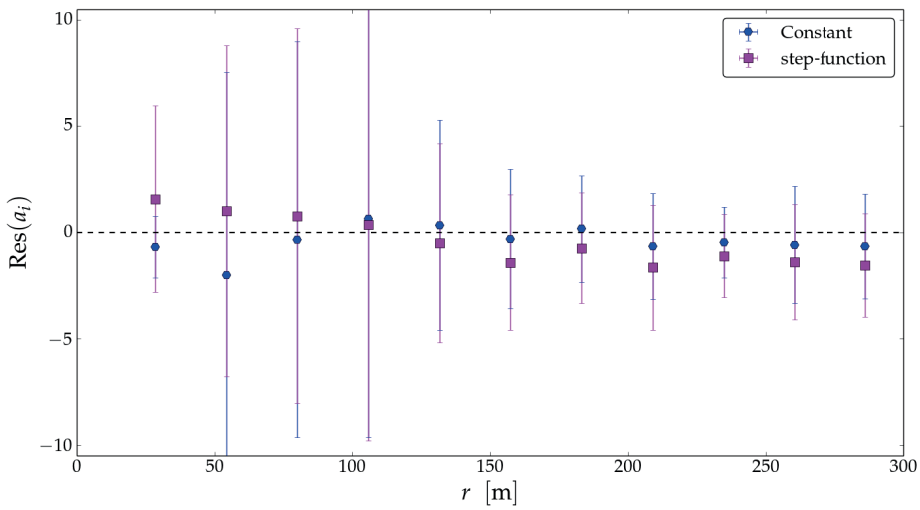

FIGURE 6.15: The mean of the residuals of the expected charge-excess fraction as in Eq. (6.15) per distance bin i, for the fitted constant and the smooth-step model, after the bootstrapping procedure to estimate the parameters.

provided by the model in Eq. 6.14, reproduces the data better than the constant approximation. Beyond 130 m, the constant approximation evidently provides the best agreement with the data.

6.3 Conclusions

In this Chapter, we performed a polarization study of the radio emission from EAS's recorded by the dense core of AERA, and externally triggered by the AERAlet SD array. Such CR events have been reconstructed by means of $\overline{\text{Off}\underline{\text{line}}}$ using the pipeline specified in Section 4.3, to extract the radio-signal polarization information. This information is exploited to estimate the fraction of the radially-polarized electric-field component, which is largely due to the charge-excess emission mechanism, relative to the dominant electric-field component polarized along the Lorentz-force direction (see Section 2.2.1).

An initial observation of this polarization signature of the radio signal was performed in the initial stages of the development of AERA, which estimated the average relative fraction of the charge-excess contribution to $14 \pm 2\%$ [54]. Recently, similar polarization measurements conducted by the LOFAR collaboration [115] have confirmed the dependency of this fraction on the observer distance with respect to the shower axis, first predicted in [45, 114]. At the shower axis, the geomagnetic contribution to the radio signal dominates, and as a consequence the fraction of charge-excess contribution vanishes.

From the selected dataset of CR events, the relative charge-excess fraction per station has been calculated using Eq. (6.5). The use of this equation is problematic for certain observer positions, namely where the geomagnetic and the charge-excess mechanisms interfere strongly. In addition, as it can be seen from Fig. 6.1, large uncertainties for the angular variables give rise to very large uncertainties for the calculated fraction. For this reason we have developed a new method to assess the uncertainty in the observer angles using a toy Monte Carlo model. This model mainly relies on the uncertainty magnitudes of the angular variables involved, which is a lot better than a priori excluding a certain range.

After that, the calculated charge-excess fraction for the selected stations has been analyzed as a function of the station distances from the shower axis, in a multi-event approach. From the data side, a small relative fraction of the charge-excess contribution is found close to the shower axis, up to a distance of 70 m far from it. Beyond this distance, around the criti-

cal distance of $\approx$ 90 m, we observe a growth and a consecutive saturation of the charge-excess fraction (cf. Fig. 6.7).

In the hypothesis that such a trend is not significant, an average charge-excess fraction can be estimated by fitting a constant trough the data, which gives back the value of $13 \pm 4\%$ in striking agreement with the result in [54], using a different data set. However, the model given by Eq. (6.11) seems to fit the data better than a constant fraction throughout the whole range of distances, though the wide spread of the data points makes the usual χ^2 comparison unusable. Indeed, it is not clear whether this spread is genuinely due to the air-shower zenith angle or reflects the instabilities of the formula in Eq. (6.5). Therefore, we have included in the model in Eq. (6.12), the additional zenith-angle dependency observed in [115]. Although a slight improvement is achieved, there is again no definitive conclusion. Finally, as a last attempt, the function Eq. (6.14) describes a (smooth) transition between two possible values of the charge excess fraction (cf. Fig. 6.8), as close to and far away from the shower axis.

For this reason, a bootstrapping approach has been followed to accurately estimate the uncertainties of the parameters for the different models. Basically, by bootstrapping, the fitting procedure is repeated many times. Therefore one can infer the sampling distributions of the model parameters, to estimate their uncertainties and compare them with the single-fit results. In the simplest case, when fitting a constant, the sampling distribution of the charge-excess fraction can be modelled by a generalized extreme value distribution (cf. Fig. 6.10). Indeed, the 95% confidence interval probability of this distribution (indicated by the black-solid lines in this Figure) is $\approx$ 0.05-0.25. These values are close to the α_0, α_1 parameter estimates resulting by bootstrapping the smooth-step model indicated in Table 6.9. At the same time, though this method rules out the model in Eq. (6.11), it gives a hint that the zenith-angle dependence needs to be included.

From a strictly-statistical point of view, at the AERA site, the radially-polarized contribution to the total radio signal does not show a significant dependency on the shower-axis distance of the observer. Though, at a qualitative level, we found that such a contribution is essentially vanishing in the region close to the shower core up to a "critical distance" where it becomes (and saturates to a value) significantly different from zero. The present sensitivity of the array has not allowed us to determine whether this change is described by a gradual or a sudden growth. The wide scatter of the measurements (with respect to their uncertainties), reflected in the large spread of the residual distributions for the different

models, makes it hard to compare between the assumptions not ruled out by the bootstrapping approach. The missing atmospheric-electric-field monitoring has influenced the event and station selection (which could explain the scatter of the data), as we have only relied on data-driven selection procedures such as the ones regarding the reliability of the polarization-angle estimates.

"[...] Perhaps a physicist would know at once why this whole idea was absurd. But then, perhaps a physicist would be so locked into the consensus of his scientific community that it would be harder for him to accept an idea that transformed the meaning of everything he knew. Even if it were true."

O. S. Card, [*XENOCIDE*, p. 65, Tor Ed.]

"In our description of nature the purpose is not to disclose the real essence of the phenomena but only to track down, so far as it is possible, relations between the manifold aspects of our experience."

N. Bohr, *Atomic Theory and the Description of Nature*

7

Summary and Discussion

With this work, we have opened a low-energy window on UHECR research at the Pierre Auger Observatory. In addition to the HEAT enhancement, AERAlet offers the opportunity of observing EASs down to $E \gtrsim 6 \cdot 10^{16}$ eV. More than locally lowering the energy threshold, it also acts as an external trigger for the AERA antennas, giving the opportunity of simultaneously measuring the induced radio signal and improving the core-position reconstruction.

The deployment of the AERAlet facility has first required the tuning of the event-reconstruction process, which ultimately provides the primary-CR energy (among other observables). In particular, by sampling the EAS particle density at the observation level, one can infer how this particle density changes with the distance from the shower axis, by means of the LDF. As the LDF slope and the (optimal) normalization factor depend on the mutual distance between the detectors, we have deduced the suitable coefficients for the slope parametrisation, as well as the convenient distance and zenith angle needed to define a reference signal acting as an energy estimator. Finally, the AERAlet energy calibration provides the conversion of this estimator to an energy measurement by the array.

This energy estimate has then been used to extend the UHECR flux spectrum down to $\sim 6 \cdot 10^{16}$ eV. From the comparison with the energy spectrum calculated by other experiments, such as HEAT, TA, and

KASCADE – Grande, we have shown that the AERAlet spectrum agrees within $\sim 1\sigma$ with the other spectra. Though we would have needed more data to reduce the uncertainty, this proves the reliability of the AERAlet event-reconstruction and energy estimation, which can be used by the Auger collaboration for future measurements. An additional (effective) year of observations would be already enough to considerably improve the uncertainties on the flux values, and to shed some light on the features found by TA and KASCADE – Grande (discussed at the end of Section 5.4.3).

As a matter of fact, the radio-detection technique is an excellent tool for calorimetric measurements of the electromagnetic component of air showers. One can monitor the behaviour of SD, which depends on ageing of its PMTs, the water quality, and reflectivity. Therefore, SD can be monitored better than by using FD, which in turn depends on atmospheric corrections and laser calibrations. For this reason, in addition to being a physics tool (extension to low energies) the combined AERAlet + AREA phase I detector is a calibration tool for the electromagnetic component of the shower and ASCII (the upgrade of the Pierre Auger Observatory in deployment phase).

The recorded radio signals by the AERA phase I antennas, have been used to map the lateral behavior of the electric-field polarization. This information is strictly related to emission mechanisms involving the dynamics of the charged particles in the shower front, and it is exploited to calculate the contribution to the total signal of the radially-polarized electric field relative to the dominant geomagnetic one. We found that this fraction is basically zero along the shower axis, where the emission is purely geomagnetic, and starts to be relevant at a lateral distance of 90-100 m. Unfortunately, we can state this only qualitatively. The uncertainties on the angular variables, which characterize the observer position and detected polarization of the signal, are further amplified when propagated down to the calculated fraction of radially polarized signal even after the proposed additional cuts in Section 6.1. These uncertainties are mainly related to the accuracy of the shower-core position, which although has been improved by the AERAlet reconstruction, remains a determinant factor in the extraction of the polarization information from the recorded radio pulses.

This lateral dependence of the radially-polarized contribution, mainly due to the (negative) charge-excess mechanism, confirms the recent results of the LOFAR experiment and attests our understanding of the radio-emission mechanisms in EASs.

The Pierre Auger Observatory is at present the largest operating UHECR experiment, and its activity has been further prolonged till 2023 thanks to the planned upgrade starting in 2016. This upgrade concerns the SD array of the Observatory, and it will consist in placing scintillator plates on top of the water-Cherenkov tanks of the array. Preliminary studies have already shown that such an additional detector allows the disentanglement of the muon- and electromagnetic-generated signals, which currently give rise to the total signal in the tank. In this way, it will be possible to discriminate muon-rich from muon-poor showers that respectively correspond to heavy and light hadron-initiated EASs. This means that the composition can be addressed at the same time through direct measurements, which basically does not rely on simulations. Actually, the composition is the missing piece of the UHECR puzzle, and in the next few years the Auger collaboration will have the chance of unraveling one of the most exciting mysteries of physics.

Moreover, in the coming years the Japanese Experiment Module - Extreme Universe Space Observatory (JEM-EUSO) is expected to be launched and installed on the International Space Station. This new type of observatory will detect UHECRs at the extreme energies, by means of the same principle exploited by the FD of Auger but with a much larger instantaneous aperture. This experiment will be of fundamental importance to cross-check the Auger results on the eventual GZK suppression of the UHECR flux spectrum.

Beside the astrophysical implications that can be derived by a deep understanding of UHECRs, discussed in Chapter 2, these cosmic messengers play a key role in other research fields, from Particle Physics to Astrobiology and Astrochemistry.

The study of collisions occurring at center-of-mass energies > 70 TeV (about a factor 7 higher than LHC) between UHECRs and the molecules of the Earth's atmosphere, may disclose some insights on high-energy hadronic interactions, which currently remain hardly understood. Other than providing proton-air cross-section measurements, such studies may provide new constraints on the physics beyond the Standard Model.

At the same time, those CRs arriving on Earth are a very small fraction of the whole population of such ionizing radiation spreading around through the interstellar medium of galaxies. The ionization and chemical state of H_2 gas regions, where the star-formation initial condition are set, is continuously altered by CRs [123]. In addition, as argued in [124], CRs are responsible for initiating chemical reactions in the midplane of protoplanetary disks.

In addition to star and planetary formation, not long after the discovery of CRs, researchers started to question the impact of cosmic radiation on living organisms and on the evolution/origin of life. The Miller–Urey experiment [125] demonstrated that lightnings have a major role in the creation of simple organic molecules, the building blocks of life. It is thought that lightnings are strictly connected to CRs, and recently this relation has been proved to be very realistic [126, 127]. Therefore, variations in the CR flux induce variations in the rate of lightnings, which can then contribute to the possible origin of life. Moreover, high-energy ionizing radiation damage the DNA molecules causing mutations [128]. Understanding the production mechanisms of UHECRs and the properties of the sources would help to better estimate CR-flux variabilities over long time scales, and their effects on terrestrial life.

A

Used configuration files for Offline modules

A.1 SD-Only Reconstruction Pipeline

```
<?xml version="1.0" encoding="iso-8859-1"?>

<!-- Sequence file for SD data reconstruction.-->
<sequenceFile xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
              xsi:noNamespaceSchemaLocation='@XMLSCHEMALOCATION@/ModuleSequence.xsd'>

  <enableTiming/>
    <moduleControl>
      <loop numTimes="unbounded">

        <module> EventFileReaderOG              </module>
        <module> EventCheckerOG                 </module>
        <module> TriggerTimeCorrection          </module>
        <module> SdCalibratorOG                 </module>
        <module> SdStationPositionCorrection    </module>
        <module> SdBadStationRejectorKG         </module>
        <module> SdSignalRecoveryKLT            </module>
        <module> SdEventSelectorOG              </module>

        <module> SdPlaneFitOG                   </module>
        <module> LDFFinderKG                    </module>
        <module> SdEventPosteriorSelectorOG     </module>

        <module> SdRecPlotterOG                 </module>
        <module> RecDataWriterNG                </module>
```

```
      </loop>
    </moduleControl>

</sequenceFile>
```

A.2 Radio-SD Reconstruction Pipeline

```
<!-- A sequence for processing AERA events -->

<sequenceFile xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
              xsi:noNamespaceSchemaLocation='@XMLSCHEMALOCATION@/ModuleSequence.xsd'>

  <enableTiming/>

  <moduleControl>

    <loop numTimes="unbounded">

      <module> EventFileReaderOG                  </module>
      <module> RdEventPreSelector                 </module>

      <module> EventCheckerOG                     </module>
      <module> SdQualityCutTaggerOG               </module>
      <module> SdPMTQualityCheckerKG              </module>
      <module> TriggerTimeCorrection              </module>
      <module> SdCalibratorOG                     </module>
      <module> SdStationPositionCorrection        </module>
      <module> SdBadStationRejectorKG             </module>
      <module> SdSignalRecoveryKLT                </module>
      <module> SdEventSelectorOG                  </module>
      <module> SdPlaneFitOG                       </module>
      <module> LDFFinderKG                        </module>
      <module> SdEventPosteriorSelectorOG         </module>

      <module> RdEventInitializer                 </module>
      <module> RdStationPositionCorrection        </module>
      <module> RdStationRejector                  </module>
      <module> RdChannelADCToVoltageConverter     </module>
      <module> RdChannelSelector                  </module>
      <module> RdChannelPedestalRemover           </module>
      <module> RdChannelResponseIncorporator      </module>
      <module> RdChannelBeaconSuppressor          </module>
      <module> RdStationTimeWindowConsolidator    </module>
      <module> RdStationTimingCalibrator          </module>
      <module> RdChannelTimeSeriesTaperer         </module>
      <module> RdChannelBandstopFilter            </module>
      <module> RdChannelUpsampler                 </module>
      <module> RdChannelRiseTimeCalculator        </module>

      <module> RdAntennaChannelToStationConverter </module>
      <module> RdStationSignalReconstructor       </module>
      <module> RdPolarizationReconstructor        </module>
      <module> RdStationQualityAssessor           </module>

      <module> RecDataWriterNG                    </module>

    </loop>
```

```
  </moduleControl>

</sequenceFile>
```

A.3 Configuration for the SdEventSelectorOG module

```
<!-- -*- XML -*- Configuration of Module SdEventSelector -->

<SdEventSelector xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                 xsi:noNamespaceSchemaLocation='@XMLSCHEMALOCATION@/SdEventSelector.xsd

  <!-- Minimal number of stations to be contained in an event to be
       processed (without accidentals, counting twins as one) -->
  <MinNumberOfStationsAtBeginning> 3 </MinNumberOfStationsAtBeginning>

  <!-- Apply BottomUp Selection
       (events that are 3TOT or 4C1) -->
  <EnableBottomUpSelection> 1 </EnableBottomUpSelection>

  <!-- Calculate T4 trigger
       (events that are 3TOT or 4C1) -->
  <EnableT4Trigger> 1 </EnableT4Trigger>

  <!-- Reject events that do not pass T4 criteria -->
  <RejectNonT4Events> 0 </RejectNonT4Events>

  <!-- Calculate T5 property
       (station with maximal signal has to have "MinNumberOfActiveStations"
        working neighbors) -->
  <EnableT5Trigger> 0 </EnableT5Trigger>

  <!-- Minimal number of active stations to accept the event
       as a T5 -->
  <MinNumberOfActiveStations> 6 </MinNumberOfActiveStations>

  <!-- Reject events that do not pass T5 criteria -->
  <RejectNonT5Events> 0 </RejectNonT5Events>

  <!-- Calculate T5Posterior propery
       (the 3 stations in an equilaterial triangle must be active/alive -->
  <EnableT5PosteriorTrigger> 1 </EnableT5PosteriorTrigger>

  <!-- Reject events that do not pass T5 criteria -->
  <RejectNonT5PosteriorEvents> 0 </RejectNonT5PosteriorEvents>

  <!-- Minimal number of active stations to accept the event
       as a T5 Posterior -->
  <MinNumberOfActiveStationsPosterior> 5 </MinNumberOfActiveStationsPosterior>

  <!-- lightning rejection:
     0 - keep all lightning stations
     1 - reject lightning stations
     2 - reject whole event when lightning detected in any of the stations -->
  <RejectLightning> 1 </RejectLightning>

  <!-- Enable event tagging based on a list of excluded periods (in separate
       xml file) -->
```

```
  <EnableSdExcludedPeriods> 1 </EnableSdExcludedPeriods>

  <!-- Reject events that fall in time periods, where the SD detector
       had serious problems. This is mainly important when you try
       to reconstruct an energy spectrum (calculating the acceptance,
       etc.) Set this to 0 to keep events within bad periods. -->
  <RejectBadPeriods> 0 </RejectBadPeriods>

  <!-- selection parameters -->
  <!-- GridType: 0 = standard, 1 = infill, 2 = infill-of-the-infill -->
  <GridType> 2 </GridType>
  <LightCompatibilityTolerance unit="nanosecond"> 65 </LightCompatibilityTolerance>
  <FirstCrownDistanceRange unit="meter"> 360 500 </FirstCrownDistanceRange>
  <SkewMaximumDistance unit="meter"> 900 </SkewMaximumDistance>
  <SeedPlaneFrontTimeResidualRange unit="nanosecond"> -300 1000 </SeedPlaneFrontTimeResidua
  <LonelyIfNoneInDistance unit="meter"> 600 </LonelyIfNoneInDistance>
  <LonelyIfOneInDistance unit="meter"> 1250 </LonelyIfOneInDistance>

  <!-- Reject stations with trigger TOTd? -->
  <RejectTOTdStations> 1 </RejectTOTdStations>
  <!-- Reject stations with trigger MOPs? -->
  <RejectMoPSStations> 1 </RejectMoPSStations>

</SdEventSelector>
```

A.4 Configuration for the `LDFFinderKG` module

```
<?xml version="1.0" encoding="iso-8859-1"?>

<!-- Configuration of Module LDFFinder -->

<LDFFinder xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
           xsi:noNamespaceSchemaLocation='@XMLSCHEMALOCATION@/LDFFinderKG.xsd'>

  <!-- verbosity 0=none, 1=final result, 2=intermediate,
                 3=most (obscure), 4=minuit included -->
  <infoLevel> 4 </infoLevel>

  <!-- optimal radius for LDF(r) = 1, depends on grid spacing -->
  <ldfReferenceDistance unit="meter"> 250 </ldfReferenceDistance>

  <!-- type of LDF model -->
  <ldfType> NKGBetaOnly </ldfType>

  <signalVariance> eGAP2014_035 </signalVariance>

  <!-- parameters of LDF shape model, for more details
       see function ComputeShape(...) -->

  <ldfParameters>
    <!-- values from Alexander Schulz, see talk in Santiago, June 2011 -->
    -2.483 -0.115 0.258 0.036 0.265 -0.070
  </ldfParameters>

  <!-- parameters of the shape uncertainty model -->
  <ldfUncertaintyParameters>
    <!-- values from Alexander Schulz, see talk in Santiago, June 2011 -->
    1.0438 -1.089
```

```
</ldfUncertaintyParameters>

<!-- Energy calibration:
   attenuation:
     x(theta) = cos^2(theta) - cos^2(38*degree)
     CIC(x) = 1 + attenuationPar1 * x + attenuationPar2 * x^2 +
                  attenuationPar3 * x^3
     S38 = S(rOpt) / CIC(x(theta))

   energy = energyS38Const * pow(S38, energyS38Slope)
-->
<energyCalibration>
  <!-- values from Hans Dembinski, see talk in Santiago, June 2011 -->
  <referenceAngle unit="degree"> 35 </referenceAngle>
  <attenuationPar1> 1.56 </attenuationPar1>
  <attenuationPar2> -1.21 </attenuationPar2>
  <attenuationPar3> 0 </attenuationPar3>

  <energyS38Const unit= "PeV"> 13.9 </energyS38Const>
  <energyS38Slope> 0.98 </energyS38Slope>
</energyCalibration>

<!-- core type:
  MC     : simulated true core
  Hybrid : best fitted hybrid geometry
  Fit    : core is fitted from SD data -->
<coreType> Fit </coreType>

<!-- whether or not to fix the shower axis -->
<fixedAxis> 0 </fixedAxis>

<!-- whether or not to perform a final global fit of LDF and shower front -->
<globalFit> 0 </globalFit>

<!-- maximal zenith Theta angle to perform LDF fit -->
<maxTheta unit="degree"> 80 </maxTheta>

<!-- minimization to use: MaxLike, Chi2 -->
<minimizationMethod> MaxLike </minimizationMethod>

<!-- maximal Chi2/ndof allowed -->
<maxChi2> 1e6 </maxChi2>

<!-- use silent stations in the fits -->
<useSilentStations> 1 </useSilentStations>

<!-- use silent stations from other grids in the fits -->
<useSilentStationsFromOtherGrids> 1 </useSilentStationsFromOtherGrids>

<!-- use saturation recovery information if available -->
<useSaturationRecovery> 1 </useSaturationRecovery>

<!-- discard signal recovery information if 2nd derivative of LDF
     at the saturated station is larger than this -->
<recoveryThreshold> 1.0 </recoveryThreshold>

<!-- use silents but only up to this distance from core -->
<silentMaxRadius unit="meter"> 10000 </silentMaxRadius>
```

```
<!-- and this with a soft transition in chi2 function -->
<silentRadiusTransition unit="meter"> 1000 </silentRadiusTransition>

<!-- varying the beta estimate RoptN times within a given
     relative uncertainty of beta -->
<RoptN> 2 </RoptN>

<!-- signal threshold used when silent stations are included in
     the chi2 fits -->
<silentSignalThreshold> 3.0 </silentSignalThreshold>

<silentStations>
  <!-- do not include in the fit silent stations if
       - the event has more than minNumberOfStations and
       - the LDF value is larger than maxLDFValue and
       - the distace to the axis of the silent is smaller than maxDistance
       - if the chi2 including the silent stations is larger than chi2Difference
  -->
  <maxLDFValue> 8 </maxLDFValue>
  <maxDistance unit="meter"> 800 </maxDistance>
  <minNumberOfStations> 5 </minNumberOfStations>
  <chi2Difference>  1e10   </chi2Difference>
</silentStations>

<!-- maximal distance between potential core and barycenter -->
<maxBaryToCoreDistance unit="meter"> 2000 </maxBaryToCoreDistance>

<stage0>

  <!-- how many stations required for this stage -->
  <minNumberOfStations> 3 </minNumberOfStations>

  <!-- use CorePosition obtained by preceeding modules, otherwise
       signal-weighted barycenter is used -->
  <useCorePosition> 1 </useCorePosition>

</stage0>

<stage1>

  <!-- how many stations required for this stage -->
  <minNumberOfStations> 3 </minNumberOfStations>

</stage1>

<stage2>

  <!-- how many stations required to fit (relax) slope beta -->
  <minNumberRelaxBeta> 4 </minNumberRelaxBeta>

  <!-- how many stations required to fit (relax) slope gamma -->
  <minNumberRelaxGamma> 5000 </minNumberRelaxGamma>

</stage2>

<!-- Shower curvature -->
<stage3>

  <!-- how many stations required to fit curvature radius and axis?
```

```
        note that for less stations a parameterized curvature will
        be used -->
    <minNumberOfStations> 3 </minNumberOfStations>

    <!-- how many stations for unparameterized curvature fit ? -->
    <minNumberForFullCurvatureFit> 5 </minNumberForFullCurvatureFit>

    <!-- maximum allowed change of shower axis (in solid angle difference) -->
    <maxAxisAngleDifference unit="degree"> 15 </maxAxisAngleDifference>

    <!-- reject time outliers by performing multiple fits of the curvature
        by rejecting one station at a time and by checking that the chi2
        probability becomes 2 times better -->
    <outliersRejection> 0 </outliersRejection>

  </stage3>

</LDFFinder>
```

B

Estimation of the Variance for normally-distributed angular variables on the unit circle

In Chapter 6, and in particular in Section 6.1.2, we used a Monte Carlo method to estimate the distribution of the angular variable ϕ. Hence, such a distribution has support on the unit circle S^1, and it results that can be approximated by the wrapped normal distribution around S^1. As an example, Fig. B.1 shows one of the distributions resulting from the aforementioned Monte Carlo method, with $\phi = 30^\circ$. The usual sample variance estimation for this distribution returns $\sim 167^\circ$, which is clearly wrong.

B.1 Variance Estimation

A close approximation to the wrapped normal distribution is the von Mises distribution, a continuous probability distribution on the unit circle [129, 130]:

$$f(x|\mu,\kappa) = \frac{e^{\kappa\cos(x-\mu)}}{2\pi I_0(\kappa)}, \qquad \text{with} \qquad \mu,\kappa > 0 \in \mathbb{R} \tag{B.1}$$

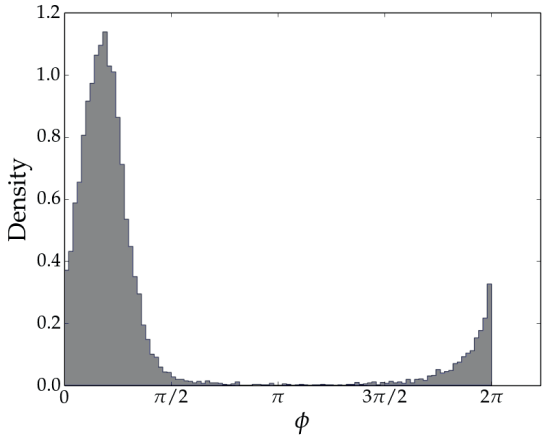

FIGURE B.1: Example of a wrapped normal distribution generated by the Monte Carlo method described in Section 6.1.2, with $\phi = 30^\circ$.

with x in any interval of length 2π, and $I_0(\kappa)$ the modified Bessel function of order 0. For this p.d.f. the location parameter μ plays the role of the mean, while the reciprocal of the concentration parameter κ the role of the variance of a normal p.d.f. though $1/\kappa$ converges to the variance for large κ.

From the moments calculation of the Von Mises distribution, it is possible to see that the "circular variance" of x is given by [131]:

$$\mathrm{Var}(x) = 1 - \frac{I_1(\kappa)}{I_0(\kappa)}, \tag{B.2}$$

where $I_1(\kappa)$ is the modified Bessel function of order 1. A simple approximation to κ is [132]:

$$\begin{aligned} \hat{\kappa} &= \frac{\bar{R}(2-\bar{R}^2)}{1-\bar{R}^2} \qquad \text{with} \\ \bar{R}^2 &= \left(\frac{1}{N}\sum_{n=1}^{N}\cos\phi_n\right)^2 + \left(\frac{1}{N}\sum_{n=1}^{N}\sin\phi_n\right)^2, \end{aligned} \tag{B.3}$$

which can be further improved by the Newton–Raphson (recursive) method.

Implementing the steps above, we used the estimated κ to evaluate the ratio of the Bessel's functions. Finally replacing this ratio in Eq. (B.2), we obtained an estimate for the (sample) circular variance. For the specific example above, we get $\mathrm{Var}(\phi) \simeq 6.14^\circ$.

List of Publications

Internal Publications

- A. M. van den Berg, D. Fraenkel, S. Messina, D. M. Varnav, F. Contreras, R. Sato, G. Zarza. *Fiber Communication System for the 433 m AERAlet SD infill of the Auger Engineering Radio Array at the Pierre Auger Observatory*, 2012.

- S. Messina, A. M. van den Berg. *Reconstruction comparison between Infill and AERAlet array*, 2013.

- S. Messina, A. M. van den Berg. *Reconstruction of 'AERA'–triggered air showers*, 2013.

- S. Messina, J. van den Eijnden, A. M. van den Berg. *Energy Calibration for the 433 m SD Infill Array*, 2014.

Publications by the Pierre Auger Observatory

2012

- Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., Allard, D., ..., Martin, L. *Large-scale distribution of arrival directions of cosmic rays detected above* 10^{18} eV *at the Pierre Auger Observatory*. Astrophysical journal supplement series, 203(2), doi:10.1088/0067-0049/203/2/34

- Charrier, D., Garçon, T., Rivière, C., Stassi, P., Pierre Auger Collaboration, . N. V., Abreu, P., ..., Acounis, S. *Results of a self-triggered*

prototype system for radio-detection of extensive air showers at the Pierre Auger Observatory. Journal of Instrumentation, 7(11), 11023.

- Settimo, M., Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., . . . , Martin, L. *Measurement of the cosmic ray energy spectrum using hybrid events of the Pierre Auger Observatory*. European physical journal plus, 127(8). doi:10.1140/epjp/i2012-12087-9

2013

- Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., Allard, D., . . . , Martin, L. *Constraints on the origin of cosmic rays above* 10^{18} eV *from large-scale anisotropy searches in data of the Pierre Auger Observatory*. Astrophysical Journal Letters, 762(1), doi:10.1088/2041-8205/762/1/L13

- Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., Allekotte, I., . . . , Martin, L. *Identifying clouds over the Pierre Auger Observatory using infrared satellite data*. Astroparticle Physics, 50-52, 92-101.

- Letessier-Selvon, A., for the Pierre Auger Collaboration, . N. V., :, . N. V., Aab, A., Abreu, P., Aglietta, M., ... Martin, L. *Highlights from the Pierre Auger Observatory*. [arXiv 1310.4620]

- Abu-Zayyad, T., Allen, M., Anderson, R., Azuma, R., . . . , Martin, L. *Pierre Auger Observatory and Telescope Array: Joint Contributions to the 33rd International Cosmic Ray Conference*. [arXiv 1310.647]

- Aab, A., Abreu, P., Aglietta, M., Ahlers, M., Ahn, E-J., . . . , Ziolkowski, M. *The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference*.

- Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., Allekotte, I., . . . , Martin, L. *Bounds on the density of sources of ultra-high energy cosmic rays from the Pierre Auger Observatory*. Journal of Cosmology and Astroparticle Physics, (5). doi:10.1088/1475-7516/2013/05/009

- Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., Allekotte, I., ..., Martin, L.. *Techniques for measuring aerosol attenuation using the Central Laser Facility at the Pierre Auger Observatory*. Journal of Instrumentation, 8. doi:10.1088/1748-0221/8/04/P04009

- Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., Allekotte, I., ..., Martin, L. *Interpretation of the depths of maximum of extensive air showers measured by the Pierre Auger Observatory*. Journal of Cosmology and Astroparticle Physics, (2). doi:10.1088/1475-7516/2013/02/026

2014

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albuquerque, I. F. M., ...Ziolkowski, M.*Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above* $10^{17.8}$ eV. Physical Review D, 90 (12). doi:10.1103/PhysRevD.90.122005

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albuquerque, I. F. M., ..., Ziolkowski, M.*Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications*. Physical Review D, 90 (12). 10.1103/PhysRevD.90.122006

- Aab, A., Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., ..., Ziolkowski, M. *Origin of atmospheric aerosols at the Pierre Auger Observatory using studies of air mass trajectories in South America*. Atmospheric research, 149, 120-135. doi:10.1016/j.atmosres.2014.05.021

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albuquerque, I. F. M., ..., the Telescope Array Collaboration. *Searches for large-scale anisotropy in the arrival directions of cosmic rays detected above energy* 10^{19} eV *at the Pierre Auger Observatory and the Telescope Array*. The Astrophysical Journal, 794 (2). doi:10.1088/0004-637X/794/2/172

- Aab, A., Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., ..., Ziolkowski, M. *Reconstruction of inclined air showers detected with the Pierre Auger Observatory*. Journal of Cosmology and Astroparticle Physics, 08. doi:10.1088/1475-7516/2014/08/019

- Aab, A., Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Al Samarai, I., ..., Ziolkowski, M. *Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth*. Physical Review D, 90 (1). doi:10.1103/PhysRevD.90.012012

- Aab, A., Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Samarai, I. A., ..., Ziolkowski, M. *A Search for Point Sources of* EeV *Photons*. The Astrophysical Journal, 789(2), 160. 10.1088/0004-637X/789/2/160

- Aab, A., Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., ..., Ziolkowski, M. *A Targeted Search for Point Sources of EeV Neutrons*. Astrophysical Journal Letters, 789 (2), 4038. doi:10.1088/2041-8205/789/2/L34

- Aab, A., Abreu, P., Aglietta, M., Ahlers, M., Ahn, E. J., Albuquerque, I. F. M., ..., Martin, L. *Probing the radio emission from air showers with polarization measurements*. Physical Review D, 89, 1-18. doi:10.1103/PhysRevD.89.052002

2015

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albert, J. N., ..., Zuccarello, F. *The Pierre Auger Cosmic Ray Observatory*. Nuclear instruments & methods in physics research section A - Accelerators spectrometers detectors and associated equipment, 798, 172-213. doi:10.1016/j.nima.2015.06.058

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albuquerque, I. F. M., ..., Zuccarello F. *Measurement of the cosmic ray spectrum above 4 x 10(18) eV using inclined events detected with the Pierre Auger Observatory*. Journal of Cosmology and Astroparticle Physics, (8). doi:10.1088/1475-7516/2015/08/049

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albuquerque, I. F. M., ..., Zuccarello F. *Erratum: Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth*. Physical Review D, 92 (1).
- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Samarai, I. A., Albuquerque, I. F. M., ..., Zuccarello F. *Search for patterns by combining*

cosmic-ray energy and arrival directions at the Pierre Auger Observatory. European Physical Journal C, 75 (6). doi:10.1140/epjc/s10052-015-3471-0

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albuquerque, I. F. M., . . . , Zuccarello F. *Improved limit to the diffuse flux of ultrahigh energy neutrinos from the Pierre Auger Observatory*. Physical Review D, 91 (9). doi:10.1103/PhysRevD.91.092008

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Samarai, I. A., Albuquerque, I. F. M., . . . , Zuccarello, F. *Large Scale Distribution of Ultra High Energy Cosmic Rays Detected at the Pierre Auger Observatory with Zenith Angles up to* 80°. The Astrophysical Journal, 802, 111.

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Al Samarai, I., Albuquerque, I. F. M., . . . , Zuccarello F. *Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events*. Physical Review D, 91.

- Aab, A., Abreu, P., Aglietta, M., Ahn, E. J., Samarai, I. A., Albuquerque, I. F. M., . . . , Zuccarello, F. *Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory*. The Astrophysical Journal, 804, 15.

Acronyms

a.s.l. above sea level. 20
ADST Advanced Data Summary Tree. 46
AERA Auger Engineering Radio Array. 5
AGN Active Galactic Nuclei. 11
AMIGA Auger Muons and Infill for the Ground Array. 5
ASCII Auger Scintillators for Composition II. 5

CDAS Central Data Acquisition System. 28
CIC Constant Intensity Cut. 22
CMBR Cosmic Microwave Background Radiation. 2
COAST COrsika dAta accesS Tools. 17
CORSIKA COsmic Ray SImulations for KAscade. 16
CR Cosmic Ray. 2
CRS Central Radio Station. 36

DAQ Data AcQuisition system. 36

EAS Extensive Air Shower. 4
EBL Extragalactic Background Light. 2
EGCR Extra-Galactic Cosmic Ray. 8

FADC Flash Analog to Digital Converter. 31
FD Fluorescence Detector. 27

GCR Galactic Cosmic Ray. 8
GPS Global Positioning System. 30
GRB Gamma-Ray Burst. 11

HEAT High-Elevation Auger Telescopes. 5

IR Infra-Red. 2

KASCADE KArlsruhe Shower Core and Array DEtector. 94

LDF Lateral (density) Distribution Function. 18
LPDA Log-Periodic Dipole Antenna. 35
LS Local Station. 37

MC Monte Carlo. 50

NKG Nishimura-Kamata-Greisen. 44

p.d.f. probability density function. 92
PAO Pierre Auger Observatory. 4, 5
PLD Programmable Logic Device. 31
PMT Photo-Multiplier Tube. 30

RDS Radio-Detection Station. 35
RFI Radio-Frequency Interference. 48
RMS Root Mean Square. 48

SD Surface Detector. 27

TA Telescope Array. 93

UHECR Ultra-High Energy Cosmic Ray. 1, 3

VCT Vertical and Central Through-going. 31
VEM Vertical Equivalent Muon. 31

WCD Water-Cherenkov Detector. 30
WLAN Wireless Local Area Network. 29

List of Tables

5.1 Summary of the number of remaining events after each applied cut 70
5.2 Summary of the optimal distances for different array spacings 73
5.3 Parameters for the LDF-slope model 75
5.4 Parameters for the LDF-slope relative-uncertainty model . . 76
5.5 Parameters for the LDF-slope model after the anti-bias cut . 80
5.6 Parameters for the constant-intensity cut correction 89
5.7 Energy calibration parameters 101

6.1 Conjunction of the possible values of charge-excess fraction, polarization angle, and observer angle 122
6.2 Observer-angle Uncertainty model parameters 125
6.3 Fitted parameters for the *r-only* model 132
6.4 Fitted parameters for the *r-θ* model 132
6.5 Fitted parameters for the simple *r-θ* model 133
6.6 Fitted parameters for the smooth-step model 134
6.7 Weighted averages of the *r-only* model parameters after the bootstrap 138
6.8 Weighted averages of the *r-θ* model parameters after the bootstrap 138
6.9 Fitted parameters for the smooth-step model after the bootstrap 139

List of Figures

1.1 Spectral energy densities in intergalactic space of various radiations . . . 2
1.2 Sky map of cosmic-ray arrival directions . . . 4

2.1 The cosmic-ray energy spectrum . . . 10
2.2 Model predictions for the cosmic-ray-flux dip region . . . 12
2.3 The Hillas diagram . . . 16
2.4 Schematic representation of a developing cascade . . . 19
2.5 Schematic representation of a developing air shower . . . 21
2.6 Illustrative example of different-primary-generated air showers . . . 22
2.7 Example of Longitudinal Profiles . . . 23
2.8 Example of shower-to-shower fluctuations . . . 24
2.9 Example of two-dimensional LDF . . . 26
2.10 Particle-density at the ground for three different zenith angles 27
2.11 Illustration of the diffrent radio-emission contributions . . . 29
2.12 Simulated footprints of the radio signal with CoREAS . . . 32

3.1 Layout of the Pierre Auger Observatory . . . 37
3.2 Scheme of an FD building, and of an FD telescope . . . 38
3.3 Picture and Scheme of an SD local station . . . 40
3.4 Example of charge and peak histograms for a single station 41
3.5 Illustration of the 3ToT configurations . . . 44
3.6 Illustration of the 4C1 configurations . . . 44
3.7 Layout of AERA . . . 45
3.8 Picture of AERA stations . . . 47
3.9 Layout of the AERAlet array . . . 48

4.1 Illustration of the posterior T5 condition . . . 54
4.2 Schematic illustration of the planar- and spherical-front approximations . . . 55

4.3 Example of a reconstructed CR event by Offline, as wiew in the Event Browser 59
4.4 The chosen reference frame 62

5.1 T2 uptime of the seven AERAlet stations during August 2013 66
5.2 Coincidence rate versus station multiplicity 67
5.3 r_{opt} distribution of reconstructed events with fitted LDF slope 71
5.4 Shower-core positions of the reconstructed events with fitted LDF slope 72
5.5 r_{opt} as a function of the maximum-signal-station distance. . . 73
5.6 Comparison of the slope-uncertainty model with the data . 77
5.7 Distribution of the residuals for the signal model 78
5.8 Correction factor versus reference signal before and after the anti-bias cut 79
5.9 Distribution of the residuals for the signal model 80
5.10 Signal-model residuals as a function of radial distances from the core 82
5.11 Intensity of events with S_{270} above S_{270}^{cut} 85
5.12 The correction function corresponding to a *CIC* at 200 events, and the corresponding intensity spectra 87
5.13 Variation of the CIC parameters 88
5.14 Consistency of the CIC assumption 89
5.15 Variation of the correction function 91
5.16 Geometrical-reconstruction comparison between AERAlet and the 750 m SD Infill array 93
5.17 Energy distribution 94
5.18 Schematic illustration of how an orthogonal cut line might introduce a bias 95
5.19 Comparison of the models of energy and reference signal relative uncertainties (E_{Infill}, and S_{33}) against the data 98
5.20 Result of the Monte Carlo method to infer the slope of the cut line 100
5.21 Energy calibration plot 101
5.22 AERAlet energy resolution 103
5.23 Two-dimensional distribution of the relative differnce between the AERAlet energy estimate and the reference energy as a fucntion of the latter 103
5.24 Two-dimensional energy–$\cos^2(\theta)$ event distribution 104
5.25 Relative acceptance for AERAlet 106
5.26 Sketch of the AERAlet elementary surface element 106
5.27 AERAlet exposure as a fucntion of time 108

5.28 Daily-rate distribution of T5 events 109
5.29 Zenith-angle (efficiency) thresholds superimposed to the $\cos^2(\theta)$–energy event distribution 110
5.30 Cosmic-ray flux obtained with AERAlet measurements . . . 112
5.31 Comparison of the energy spectrum with other experiments 115
5.32 Comparison of the energy spectrum with KASCADE – Grande 116

6.1 Charge-excess fraction as a function of the polarization and observer angles 122
6.2 Schematic illustration of the influence of the core-location uncertainty on the determination of the observer angle. . . . 124
6.3 Estimated observer-angle uncertainty versus core-station distance 125
6.4 Examples of observer-angle cosine distribution 127
6.5 Observer-angle and polarization-angle uncertainty distributions 128
6.6 The angle of polarization as a function of the observer position in the shower plane 129
6.7 Charge-excess fraction as a function of the station distance from the showr axis 131
6.8 Comparison between data and the proposed models 133
6.9 Residuals for the expected charge-excess fraction as a function of the radial distance 135
6.10 Distribution of the fitted constants 137
6.11 Distributions of the *r-only* model parameters 137
6.12 Distributions of the *r-θ* model parameters 139
6.13 Distributions of the Smooth-Step model parameters 140
6.14 Comparison between data and the proposed models afrer the bootstrap 141
6.15 Residuals for the expected charge-excess fraction as a function of the radial distance after bootstrapping 141

B.1 Example of a wrapped normal distribution 158

Bibliography

[1] C.D. Dermer and G. Menon, *High Energy Radiation from Black Holes: Gamma Rays, Cosmic Rays, and Neutrinos* , Princeton Series in Astrophysics. Princeton University Press, 2009.

[2] Eli Waxman and John N. Bahcall, "High-energy neutrinos from astrophysical sources: An Upper bound", *Phys. Rev.*, vol. D59, pp. 023002, 1999.

[3] Alexander Aab et al., "Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory", 2014.

[4] Shoichi Ogio, "The Telescope Array Low Energy Extension", in *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, p. 0717.

[5] V.F. Hess, "ber die Beobachtungen der durchdringenden Strahlung bei sieben Freiballonflgen", *Phys. Z.*, vol. 8, pp. 1084, 1912.

[6] C. D. Anderson, "The Positive Electron", *Phys. Rev.*, vol. 43, pp. 491–494, 1933.

[7] J. C. Street and E. C. Stevenson, "New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron", *Phys. Rev.*, vol. 52, pp. 1003–1004, 1937.

[8] G. P. S. Occhialini and C. F. Powell, "Nuclear disintegration produced by slow charged particles of small mass", 1987, [Nature159,186(1947)].

[9] Pierre Auger, P. Ehrenfest, R. Maze, J. Daudin, and A. Freon Robley, "Extensive cosmic ray showers", *Rev. Mod. Phys.*, vol. 11, pp. 288–291, 1939.

[10] John Linsley, "Evidence for a primary cosmic-ray particle with energy 10**20-eV", *Phys. Rev. Lett.*, vol. 10, pp. 146–148, 1963.

[11] W. Galbraith and J. V. Jelley, "Light Pulses from the Night Sky associated with Cosmic Rays", *Nature*, vol. 171, no. 4347, pp. 349–350,

02 1953.

[12] J. V. Jelley, J. H. Fruin, N. A. Porter, T. C. Weekes, F. G. Smith, and R. A. Porter, "Radio Pulses from Extensive Cosmic-Ray Air Showers", *Nature*, vol. 205.

[13] D.M. Edge, A.C. Evans, H.J. Garmston, R.J.O. Reid, A.A. Watson, et al., "The cosmic ray spectrum at energies above 10-to-the-17 eV", *J. Phys.*, vol. A6, pp. 1612–1634, 1973.

[14] N. Chiba, K. Hashimoto, N. Hayashida, K. Honda, M. Honda, et al., "Akeno giant air shower array (AGASA) covering 100-km**2 area", *Nucl. Instrum. Meth.*, vol. A311, pp. 338–349, 1992.

[15] University of Utah, "The Fly's Eye (1981-1993)", http://www.cosmic-ray.org/reading/flyseye.html, 1981.

[16] John N. Matthews and C. C. H. Jui, "First results from the high resolution Fly's Eye experiment", *Nucl. Phys. Proc. Suppl.*, vol. 87, pp. 411–413, 2000.

[17] A. Horneffer et al., "LOPES: Detecting radio emission from cosmic ray air showers", in *SPIE Astronomical Telescopes and Instrumentation Symposium: The Industrial Revolution in Astronomy Glasgow, Scotland, United Kingdom, June 21-25, 2004*, 2004.

[18] Daniel Ardouin et al., "Radiodetection of cosmic ray extensive air showers: Present status of the Codalema experiment", *Int. J. Mod. Phys.*, vol. A20, pp. 6869–6871, 2005.

[19] IceCube Collaboration, "Measurement of the high-energy cosmic ray spectrum with IceTop", http://icecube.wisc.edu/news/view/141, 2013.

[20] Kenneth Greisen, "End to the cosmic ray spectrum?", *Phys. Rev. Lett.*, vol. 16, pp. 748–750, 1966.

[21] G.T. Zatsepin and V.A. Kuzmin, "Upper limit of the spectrum of cosmic rays", *JETP Lett.*, vol. 4, pp. 78–80, 1966.

[22] R. Aloisio, V. Berezinsky, Pasquale Blasi, A. Gazizov, S. Grigorieva, et al., "A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays", *Astropart. Phys.*, vol. 27, pp. 76–91, 2007.

[23] D. Allard, Etienne Parizot, E. Khan, S. Goriely, and A.V. Olinto, "UHE nuclei propagation and the interpretation of the ankle in the cosmic-ray spectrum", *Astron.Astrophys.*, vol. 443, pp. L29–L32, 2005.

[24] Daniel De Marco and Todor Stanev, "On the shape of the UHE

cosmic ray spectrum", *Phys. Rev.*, vol. D72, pp. 081301, 2005.

[25] Eric Armengaud, Gunter Sigl, Tristan Beau, and Francesco Miniati, "Crpropa: a numerical tool for the propagation of uhe cosmic rays, gamma-rays and neutrinos", *Astropart. Phys.*, vol. 28, pp. 463–471, 2007.

[26] Michael Unger, "Cosmic Rays above the Knee", *Invited talk at the 21st European Cosmic Ray Symposium, 2008, Kosice*, 2008.

[27] R. U. Abbasi et al., "First observation of the Greisen-Zatsepin-Kuzmin suppression", *Phys. Rev. Lett.*, vol. 100, pp. 101101, 2008.

[28] A. A. Watson, "Recent results from the Pierre Auger Observatory: Including comparisons with data from AGASA and HiRes", *Nucl. Instrum. Meth.*, vol. A588, pp. 221–226, 2008.

[29] J. Abraham et al., "Upper limit on the cosmic-ray photon fraction at EeV energies from the Pierre Auger Observatory", *Astropart. Phys.*, vol. 31, pp. 399–406, 2009.

[30] Enrico Fermi, "On the Origin of the Cosmic Radiation", *Phys. Rev.*, vol. 75, pp. 1169–1174, 1949.

[31] A.M. Hillas, "The Origin of Ultrahigh-Energy Cosmic Rays", *Ann. Rev. Astron. Astrophys.*, vol. 22, pp. 425–444, 1984.

[32] JEM-EUSO Collaboration, "Main Objective: Straight Line Astronomy Section", http://jemeuso.riken.jp/en/about2.html, 2013.

[33] D. Zavrtanik, "Ultra high energy cosmic rays", *Contemp. Phys.*, vol. 51, pp. 513–529, 2010.

[34] Peter K. F. Grieder, *Extensive Air Showers*, Springer-Verlag, 2011.

[35] D. Heck, "Extensive air shower simulations with CORSIKA and the influence of highenergy hadronic interaction models", pp. 252–259, 2000.

[36] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, "Cosmic ray interaction event generator SIBYLL 2.1", *Phys. Rev.*, vol. D80, pp. 094003, 2009.

[37] R. Ulrich, "COAST (COrsika dAta accesS Tools)", https://web.ikp.kit.edu/rulrich/coast.html, 2008, [Last Updated on Friday, 31 May 2013].

[38] P. Sommers, "Extensive air showers and measurement techniques", *Comptes Rendus Physique*, vol. 5, no. 4, pp. 463–472, MAY 2004.

[39] M. T. Dova, L. N. Epele, and Analisa G. Mariazzi, "The Effect of atmospheric attenuation on inclined cosmic ray air showers", *As-*

tropart. Phys., vol. 18, pp. 351–365, 2003.

[40] J. C. Arteaga-Velazquez et al., "The constant intensity cut method applied to the KASCADE-Grande muon data", *Nucl. Phys. Proc. Suppl.*, vol. 196, pp. 183–186, 2009.

[41] J. Alvarez-Muniz, R. Engel, T. K. Gaisser, J. A. Ortiz, and T. Stanev, "Atmospheric shower fluctuations and the constant intensity cut method", *Phys. Rev.*, vol. D66, pp. 123004, 2002.

[42] Tim Huege, "The renaissance of radio detection of cosmic rays", *Braz. J. Phys.*, vol. 44, pp. 520–529, 2014.

[43] O. Scholten, K. Werner, and F. Rusydi, "A Macroscopic Description of Coherent Geo-Magnetic Radiation from Cosmic Ray Air Showers", *Astropart. Phys.*, vol. 29, pp. 94–103, 2008.

[44] Klaus Werner and Olaf Scholten, "Macroscopic Treatment of Radio Emission from Cosmic Ray Air Showers based on Shower Simulations", *Astropart. Phys.*, vol. 29, pp. 393–411, 2008.

[45] Krijn D. de Vries, Ad M. van den Berg, Olaf Scholten, and Klaus Werner, "The Lateral Distribution Function of Coherent Radio Emission from Extensive Air Showers: Determining the Chemical Composition of Cosmic Rays", *Astropart. Phys.*, vol. 34, pp. 267–273, 2010.

[46] K. D. de Vries, A. M. van den Berg, O. Scholten, and K. Werner, "Coherent Cherenkov Radiation from Cosmic-Ray-Induced Air Showers", *Phys. Rev. Lett.*, vol. 107, pp. 061101, 2011.

[47] Klaus Werner, Krijn D. de Vries, and Olaf Scholten, "A Realistic Treatment of Geomagnetic Cherenkov Radiation from Cosmic Ray Air Showers", *Astropart. Phys.*, vol. 37, pp. 5–16, 2012.

[48] Krijn D. de Vries, Olaf Scholten, and Klaus Werner, "The air shower maximum probed by Cherenkov effects from radio emission", *Astropart. Phys.*, vol. 45, pp. 23–27, 2013.

[49] G. A. Askaryan, "Excess Negative Charge of an Electron-Photon Shower and its Coherent Radio Emission", *Soviet Physics, JETP*, vol. 14, pp. 441–443, 1962.

[50] H. R. Allan, "Radio Emission from Extensive Air Showers", *Prog. in Element. part. and Cos. Ray Phys.*, vol. 10, pp. 171, 1971.

[51] M. Galli, G. G. C. Palumbo, and G. Sinigaglia, "High frequency radio pulses from extensive air showers", *Conf. Proc.*, vol. C690825, pp. 737–742, 1969.

[52] H. R. Allan, R. W. Clay, and J. K. Jones, "Radio Pulses from Exten-

sive Air Showers", *Nature*, vol. 227, no. 5263, pp. 1116–&, 1970.

[53] D. Ardouin et al., "Geomagnetic origin of the radio emission from cosmic ray induced air showers observed by CODALEMA", *Astropart. Phys.*, vol. 31, pp. 192–200, 2009.

[54] Alexander Aab et al., "Probing the radio emission from air showers with polarization measurements", *Phys. Rev.*, vol. D89, no. 5, pp. 052002, 2014.

[55] Clancy W. James, Heino Falcke, Tim Huege, and Marianne Ludwig, "General description of electromagnetic radiation processes based on instantaneous charge acceleration in 'endpoints'", *Phys. Rev.*, vol. A84, pp. 056602, 2011.

[56] E. Hecht, *Optics*, Pearson education. Addison-Wesley, 2002.

[57] Marianne Ludwig and Tim Huege, "REAS3: Monte Carlo simulations of radio emission from cosmic ray air showers using an 'endpoint' formalism", *Astropart. Phys.*, vol. 34, pp. 438–446, 2011.

[58] Vincent Marin and Benoit Revenu, "Simulation of radio emission from cosmic ray air shower with SELFAS2", *Astropart. Phys.*, vol. 35, pp. 733–741, 2012.

[59] Jaime Alvarez-Muniz, Washington R. Carvalho, Jr., Matias Tueros, and Enrique Zas, "Coherent Cherenkov radio pulses from hadronic showers up to EeV energies", *Astropart. Phys.*, vol. 35, no. 6, pp. 287–299, JAN 2012.

[60] T. Huege, M. Ludwig, O. Scholten, and K. D. de Vries, "The convergence of EAS radio emission models and a detailed comparison of REAS3 and MGMR simulations", in *Proceedings of the ARENA2010 conference, Nantes, France*, Sept. 2010.

[61] T. Huege, M. Ludwig, and C. W. James, "Simulating radio emission from air showers with CoREAS", *AIP Conf. Proc.*, vol. 1535, pp. 128, 2013.

[62] P. Abreu et al., "Measurement of the Proton-Air Cross Section at $\sqrt{s}$=57 TeV with the Pierre Auger Observatory", *Phys. Rev. Lett.*, vol. 109, pp. 062002, Aug 2012.

[63] P. Abreu et al., "Description of atmospheric conditions at the Pierre Auger Observatory using the Global Data Assimilation System (GDAS)", *Astropart. Phys.*, vol. 35, no. 9, pp. 591 – 607, 2012.

[64] L. Nellen et al., "The observation of a muon deficit in simulations from data of the Pierre Auger Observatory", *Journal of Physics: Conference Series*, vol. 409, no. 1, pp. 012107, 2013.

[65] A. Aab et al., "Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events", *Phys. Rev. D*, vol. 91, pp. 059901, Mar 2015.

[66] "The Pierre Auger Project Design Report", 1996.

[67] Pierre Auger Collaboration, "The Pierre Auger Cosmic Ray Observatory", *Nucl. Instrum. Meth.*, vol. A798, pp. 172–213, 2015.

[68] J. Abraham et al., "The Fluorescence Detector of the Pierre Auger Observatory", *Nucl. Instrum. Meth.*, vol. A620, pp. 227–251, 2010.

[69] C. Song, Z. Cao, B. R. Dawson, B. E. Fick, Won-Yong Lee, P. Sokolsky, and X. Zhang, "Energy estimation of UHE cosmic rays using the atmospheric fluorescence technique", *Astropart. Phys.*, vol. 14, pp. 7–13, 2000.

[70] P. Abreu et al., "The Rapid Atmospheric Monitoring System of the Pierre Auger Observatory", *JINST*, vol. 7, pp. P09001, 2012.

[71] H. O. Klages, "HEAT – Enhancement Telescopes for the Pierre Auger Southern Observatory", in *Proceedings, 30th International Cosmic Ray Conference (ICRC 2007)*, 2007, vol. 5, pp. 849–852.

[72] H. Klages, "Enhancements to the Southern Pierre Auger Observatory", *J. Phys. Conf. Ser.*, vol. 375, no. 5, pp. 052006, 2012.

[73] Nils Sven Sebastian Scharf, *The energy spectrum of cosmic rays measured with the HEAT extension at the Pierre Auger Observatory*, Ph.D. thesis, Aachen, 2013, Prüfungsjahr: 2013. - Publikationsjahr: 2014; Aachen, Techn. Hochsch., Diss., 2013.

[74] I. Allekotte et al., "The Surface Detector System of the Pierre Auger Observatory", *Nucl. Instrum. Meth.*, vol. A586, pp. 409–420, 2008.

[75] Ronald C. Shellard, "First results from the Pierre Auger Observatory", *Braz. J. Phys.*, vol. 36, pp. 1184–1193, 2006.

[76] P. Bauleo, A. Etchegoyen, J. O. Fernandez Niello, A. M. J. Ferrero, A. Filevich, C. K. Guerard, F. Hasenbalg, M. A. Mostafa, D. Ravignani, and J. Rodriguez Martino, "A water tank Cherenkov detector for very high-energy astroparticles", *Nucl. Instrum. Meth.*, vol. A406, pp. 69–77, 1998.

[77] M. Aglietta et al., "Response of the Pierre Auger Observatory water Cherenkov detectors to muons", in *29th International Cosmic Ray Conference (ICRC 2005) Pune, India, August 3-11, 2005*, 2005.

[78] D. Veberič, Roth M., and H. Dembinski, *$\overline{\text{Off}}\underline{\text{line}}$ Reference Manual SD Reconstruction*, Based on Auger Internal Note GAP2005_035.

[79] H. Klages and Pierre Auger Collaboration, "Enhancements to the

Southern Pierre Auger Observatory", July 2012, vol. 375, p. 052006.

[80] Alexander Aab et al., "The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013)", in *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013)*, 2013.

[81] Pedro Abreu et al., "Antennas for the Detection of Radio Emission Pulses from Cosmic-Ray induced air showers at the Pierre Auger Observatory", *JINST*, vol. 7, pp. P10011, 2012.

[82] D. Charrier and CODALEMA Collaboration, "Antenna development for astroparticle and radioastronomy experiments", *Nuclear Instruments and Methods in Physics Research A*, vol. 662, pp. 142, Jan. 2012.

[83] S. Argiro, S.L.C. Barroso, J. Gonzalez, L. Nellen, Thomas Cantzon Paul, et al., "The Offline Software Framework of the Pierre Auger Observatory", *Nucl. Instrum. Meth.*, vol. A580, pp. 1485–1496, 2007.

[84] I. C. Mariș, "Measurement of the Ultra High Energy Cosmic Ray Flux using Data of the Pierre Auger Observatory", Auger Internal Note GAP2008_026, 2008.

[85] M. Horvata and D. Veberič, "On Shower-Front Start-Time Variance", Auger Internal Note GAP2007_057, 2007.

[86] Koichi Kamata and Jun Nishimura, "The Lateral and the Angular Structure Functions of Electron Showers", *Progress of Theoretical Physics Supplement*, vol. 6, pp. 93–155, 1958.

[87] A. Schulz, "Measurement of the Energy Spectrum of Cosmic Rays between 0.1 EeV and 30 EeV with the Infill Extension of the Surface Detector of the Pierre Auger Observatory", Auger Internal Note GAP2012_136, 2012.

[88] D. Newton, J. Knapp, and A. A. Watson, "The Optimum Distance at which to Determine the Size of a Giant Air Shower", Auger Internal Note GAP2006_045, 2006.

[89] M. Ave, P. Bauleo, John L. Harton, R. Knapik, Antonella Castellina, G. Navarra, and Aaron S. Chou, "The accuracy of signal measurement with the water Cherenkov detectors of the Pierre Auger Observatory", *Nucl. Instrum. Meth.*, vol. A578, pp. 180–184, 2007.

[90] Markus Roth, "The Lateral distribution function of shower signals in the surface detector of the Pierre Auger Observatory", in *Proceedings, 28th International Cosmic Ray Conference (ICRC 2003)*, 2003, pp. 333–336, [1,333(2003)].

[91] J. Hersil, I. Escobar, D. Scott, G. Clark, and S. Olbert, "Observations of Extensive Air Showers near the Maximum of Their Longitudinal Development", *Phys. Rev. Lett.*, vol. 6, pp. 22–23, 1961.

[92] CERN collaboration, "ROOT, the object-oriented data analysis framework", http://root.cern.ch, 2012.

[93] Eric Daniel Fraenkel, "The offline software package for analysis of radio emission from air showers at the Pierre Auger Observatory", *Nucl. Instrum. Meth.*, vol. A662, pp. S226–S229, 2012.

[94] J. Dugundji, "Envelopes and pre-envelopes of real waveforms", *Information Theory, IRE Transactions on*, vol. 4, no. 1, pp. 53–57, March 1958.

[95] Eric Daniel Fraenkel, *From Radio Pulse to Elusive Particle*, Ph.D. thesis, Rijksuniversiteit Groningen, 2014.

[96] J. P. Terlouw and M. G. R. Vogelaar, *Kapteyn Package, version 2.3*, Kapteyn Astronomical Institute, Groningen, Mar 2015, Available from http://www.astro.rug.nl/software/kapteyn/.

[97] Alexander McFarlane Mood, Franklin A. Graybill, and Duane C. Boes, *Introduction to the Theory of Statistics*, McGraw-Hill, 1963.

[98] A. J. van den Eijnden, "Statistical Analysis of the Lateral Distribution Function for the AERAlet array of the Pierre Auger Observatory", M.S. thesis, Kapteyn Institute, Rijksuniversiteit Groningen, The Netherlands, 2014.

[99] J. Abraham et al., "Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory", *Phys. Lett.*, vol. B685, pp. 239–246, 2010.

[100] S. A. M. Messina, J. van den Eijnden, and A. M. van den Berg, "Energy Calibration for the 433 m SD Infill Array", Auger Internal Note GAP2014_094, 2014.

[101] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, New York, 1993.

[102] E. Varela, "The low-energy extensions of the Pierre Auger Observatory", *J. Phys. Conf. Ser.*, vol. 468, pp. 012013, 2013.

[103] A. Hyvärinen, J. Hurri, and P.O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.*, Computational Imaging and Vision. Springer London, 2009.

[104] J. Matthews, "A Heitler model of extensive air showers", *Astropart. Phys.*, vol. 22, pp. 387–397, Jan. 2005.

[105] Design Report, "AMIGA, Auger Muons and Infill for the Ground

Array", Tech. Rep., Pierre Auger Observatory, 06 2006.

[106] Asorey H. Allekotte, I., X. Bertou, and M. Gomez Berisso, "You thought you understood hexagons?", Auger Internal Note GAP2008_114, 2008.

[107] C. Bonifazi and A. Letessier-Selvon, "Event selection using the T5 time distribution", Auger Internal Note GAP2006_042, 2006.

[108] Masaki Fukushima, "Recent Results from Telescope Array", *EPJ Web Conf.*, vol. 99, pp. 04004, 2015.

[109] C. Jui, "Summary of Results from the Telescope Array Experiment", 2015, Presented as the 34th International Cosmic Ray Conference, The Hague, The Netherlands.

[110] A. Rohatgi, "WebPlotDigitizer", http://arohatgi.info/WebPlotDigitizer.

[111] W. D. Apel et al., "Kneelike Structure in the Spectrum of the Heavy Component of Cosmic Rays Observed with KASCADE-Grande", *Phys. Rev. Lett.*, vol. 107, pp. 171104, Oct 2011.

[112] W. D. Apel et al., "Ankle-like Feature in the Energy Spectrum of Light Elements of Cosmic Rays Observed with KASCADE-Grande", *Phys. Rev.*, vol. D87, pp. 081101, 2013.

[113] W. D. Apel et al., "KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays", *Astropart. Phys.*, vol. 47, pp. 54–66, 2013.

[114] Jaime Alvarez-Muniz, Washington R. Carvalho, Jr., and Enrique Zas, "Monte Carlo simulations of radio pulses in atmospheric showers using ZHAireS", *Astropart. Phys.*, vol. 35, pp. 325–341, 2012.

[115] P. Schellart, S. Buitink, A. Corstanje, J.E. Enriquez, H. Falcke, et al., "Polarized radio emission from extensive air showers measured with LOFAR", *JCAP*, vol. 1410, no. 10, pp. 014, 2014.

[116] Harm Schoorlemmer, *Tuning in on cosmic rays. Polarization of radio signals from air showers as a probe of emission mechanisms*, Ph.D. thesis, Nijmegen U., 2012.

[117] T. W. Anderson and D. A. Darling, "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes", *Ann. Math. Statist.*, vol. 23, no. 2, pp. 193–212, 06 1952.

[118] M. A. Stephens, "EDF Statistics for Goodness of Fit and Some Comparisons", *Journal of the American Statistical Association*, vol. 69, no. 347, pp. 730–737, 1974.

[119] T. N. Gia Trinh, ", private communication, 2015.

[120] Wikipedia the free encyclopedia, "Logistic function", https://en.wikipedia.org/wiki/Logistic_function, 2015.

[121] Perry J. W., "What is the simplest formula for activation / smooth step function?", http://math.stackexchange.com/q/536041, 2013.

[122] C.Z. Mooney, R.D. Duval, and R. Duval, *Bootstrapping: A Nonparametric Approach to Statistical Inference*, Number Num. 94-95. SAGE Publications, 1993.

[123] P. P. Papadopoulos and W.-F. Thi, "The Initial Conditions of Star Formation: Cosmic Rays as the Fundamental Regulators", in *Cosmic Rays in Star-Forming Environments*, D. F. Torres and O. Reimer, Eds., 2013, vol. 34 of *Astrophysics and Space Science Proceedings*, p. 41.

[124] G.Chaparro Molano and I. Kamp, "The Cosmic-Ray Dominated Region of Protoplanetary Disks", in *Cosmic Rays in Star-Forming Environments*, Diego F. Torres and Olaf Reimer, Eds., vol. 34 of *Astrophysics and Space Science Proceedings*, pp. 133–140. Springer Berlin Heidelberg, 2013.

[125] Wikipedia the free encyclopedia, "Miller–Urey experiment", https://en.wikipedia.org/wiki/Miller%E2%80%93Urey_experiment, 2015.

[126] P. Schellart et al., "Probing Atmospheric Electric Fields in Thunderstorms through Radio Emission from Cosmic-Ray-Induced Air Showers", *Phys. Rev. Lett.*, vol. 114, no. 16, pp. 165001, 2015.

[127] Anna Dubinova, Casper Rutjes, Ute Ebert, Stijn Buitink, Olaf Scholten, and Gia Thi Ngoc Trinh, "Prediction of Lightning Inception by Large Ice Particles and Extensive Air Showers", *Phys. Rev. Lett.*, vol. 115, pp. 015002, Jun 2015.

[128] D. Atri and A. L. Melott, "Cosmic rays and terrestrial life: A brief review", *Astroparticle Physics*, vol. 53, pp. 186–190, Jan. 2014.

[129] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, Applied mathematics series. Dover Publications, 1964.

[130] Kanti V. Mardia and Peter E. Jupp, *Directional Statistics*, Wiley, 1999.

[131] Wikipedia the free encyclopedia, "von Mises distribution", 2015.

[132] Wikipedia the free encyclopedia, "Von Mises-Fisher distribution", 2015.

Samenvatting

De oorsprong van het heelal is waarschijnlijk een van de eerste vragen die is opgekomen in de gedachten van de mens, toen ze eenmaal naar de hemel begonnen te kijken met het blote oog. In feite, in het *begrijpen* van het heelal, haar oorsprong en alle gerelateerde processen, hebben astronomische waarnemingen vanaf het begin een belangrijke rol gespeeld. Van alle verschillende observatietechnieken, is de detectie van *kosmische diffuse straling* een uitzonderlijke wetenschappelijke ervaring, die ons een uitgebreid beeld geeft van het heelal voorbij het venster van zichtbaar licht.

Hier verwijst 'kosmische diffuse straling' in brede zin naar de verschillende stralingsvelden en deeltjes die ons universum opvullen. In feite wordt de gehele *ruimte* doorzeeft met diverse diffuse elektromagnetische straling en deeltjes, die de eigenschap delen dat zij *kosmisch* zijn. Dergelijke straling en deeltjes bieden ons een "glimp" welke astrofysische processen de evolutie van het heelal sturen en gestuurd hebben, en de processen, die structuur aan het heelal hebben gegeven. Met andere woorden, de aard en de geschiedenis van het heelal zijn "gecodeerd" in deze stralingsvelden en deeltjes; derhalve moet elke poging om een begrip te krijgen van de oorsprong en van de evolutionaire processen van het heelal een goede beschrijving geven van deze "stralingsvelden".

Er zijn vele bijdragen aan deze diffuse gloed, elk gekenmerkt door een zeker energiebereik. Bij de laagste energieën, onder het meV-regime, domineren radio achtergrond en de kosmische microgolf achtergrondstraling (Cosmic Microwave Background, CMB). Infrarood (IR) achtergrond, optisch licht, gammastralen, en röntgenstralen maken allemaal deel uit van het zogenaamde Extragalactische Achtergrondlicht (Extragalactic Background Light, EBL), van enkele meV tot honderden GeV. Bij de hoogste energieën, tot honderden EeV, is straling bestaande uit deeltjes, dat wil zeggen kosmische stralen (Cosmic Rays, CRs) en kosmogene neutrino's, de enige diffuse straling waar we de bronnen niet van kennen.

Elk van deze bijdragen komt noodzakelijkerwijs overeen met een specifiek proces. De CMB is in verband gebracht met de eerste processen van structuurvorming, terwijl infrarood licht voornamelijk te wijten is aan stellair licht dat verstrooid is door koude stofdeeltjes in het interstellaire medium. Optisch licht komt vooral van stellaire straling en bijdragen aan de röntgenstraling komen van stervorming en accretieschijven rond compacte sterren. Tot slot dragen zeer hoge energie gammastralen, CRs en neutrinos zeer waardevolle informatie over niet-thermische processen die "ergens" in het heelal plaatsvinden.

Sterker nog, het bestaan van dergelijke hoog-energetische deeltjes, die de atmosfeer van de aarde vanuit de ruimte binnendringen, is de boodschapper van niet-thermische verschijnselen in het heelal. De oorsprong in het universum van de hoogst-energetische deeltjes, bekend als Ultra High Energy Cosmic Rays (UHECRs), is een van de meest prominente problemen in de moderne astrofysica. In feite zijn de aankomstrichtingen van UHECRs consistent met de isotropische hypothese, en hoewel verondersteld wordt dat er slechts enkele astrofysische bronnen zijn die deeltjes naar extreme energieën kunnen versnellen, is er nog geen duidelijke correlatie met de ruimtelijk verdeling van de lichtgevende massa in het heelal (extragalactische sterrenstelsels) gevonden. Daarnaast wordt de overgang van galactische naar extragalactische kosmische straling verondersteld plaats te vinden in de *dip* van het flux spectrum, d.w.z. tussen 10^{17} en 10^{19} eV. Bij welke energie deze overgang plaatsvindt, en de afhankelijkheid ervan op de samenstelling (lading) van kosmische deeltjes, wordt gezien als een belangrijke kwestie in in UHECR onderzoek.

In dit bereik van energieën zijn verschillende afwijkingen in het flux spectrum (steiler of afvlakkend) waar te nemen ten opzichte van de gemiddelde helling van dit spectrum: spectrale kenmerken genoemd. Deze veranderingen kunnen gecorreleerd zijn met de overgang van galactische naar extragalactische bronnen of met een verandering van de chemische samenstelling (of waarschijnlijk beide) van de versnelde deeltjes. In feite zijn sommige van deze kenmerken waar te nemen rond de ondergrens van 10^{17} eV en slechts matig bestudeerd. De huidige energiedrempel van het Pierre Auger Observatorium ligt op $3 \cdot 10^{17}$ eV. Om een duidelijk beeld te krijgen van het energiebereik waar de boven genoemde overgang verwacht wordt, is het noodzakelijk om deze drempel verlagen. De uitbreiding van het energievenster voor kosmische straling in het Observatorium heeft daarom hoge prioriteit en de huidige studie vormt de basis voor de realisatie van dit doel.

Dit werk verlaagt het energie venster van UHECR onderzoek aan het

Pierre Auger Observatorium door het installeren en gebruiken van AERAlet; een deeltjes detector bestaande uit zeven stations met een onderlinge afstand van 433 m. In aanvulling op de High-Elevation Auger Telescopes (HEAT) biedt AERAlet de mogelijkheid om kosmische stralen te observeren met energieën zo laag als $E \gtrsim 6 \cdot 10^{16}$ eV. Naast het lokaal verlagen van de energiedrempel, fungeert AERAlet ook als een externe trigger voor de Auger Engineering Radio Array (AERA), waardoor de mogelijkheid ontstaat om gelijktijdig het opgewekte radiosignaal te meten en om de reconstructie van de positie met hoogste deeltjes dichtheid aan de grond te verbeteren.

Voor de inzet van de AERAlet faciliteit was het eerst nodig om de event reconstructie te optimaliseren, die uiteindelijk een schatting geeft van de primaire energie van de gedetecteerde kosmische stralen (naast andere waarneembare grootheden). Deze optimalisatie was met name bedoeld om een nieuwe parametrisatie te vinden voor de functie die gebruikt wordt om op waarnemingsniveau de laterale deeltjes dichtheid te beschrijven. Deze functie wordt de laterale verdelingsfunctie (in het Engels LDF) genoemd. Aangezien de LDF-helling en de (optimale) normalisatiefactor afhankelijk zijn van de onderlinge afstand tussen de deeltjes detectoren, hebben we geschikte coëfficiënten voor de hellingsparametrisering afgeleid, evenals de geschikte afstand en zenithoek, die nodig zijn om een referentiesignaal te definiëren dat als schatting voor de primaire energie fungeert.

Deze energieschatting is vervolgens gebruikt om het UHECR fluxspectrum uit te breiden naar $\sim 6 \cdot 10^{16}$ eV. Uit een vergelijking met het energiespectrum, berekend door andere experimenten, zoals HEAT, Telescope Array (TA) en KArlsruhe Shower Core and Array DEtector (KASCADE) – Grande, hebben we laten zien dat het AERAlet-spectrum overeenstemt binnen één standaardafwijking met de andere spectra. Hoewel we meer gegevens nodig hebben om de onzekerheid te verminderen, bewijst dit de betrouwbaarheid van de AERAlet-event reconstructie en energieschatting, die kan worden gebruikt door de Auger-collaboratie voor toekomstige metingen. Een extra (effectief) jaar van waarnemingen zou al genoeg zijn om een aanzienlijke verbetering in de onzekerheden van de fluxwaarden te leveren, en om meer licht te werpen op de kenmerken die gevonden zijn door TA en KASCADE – Grande.

In feite, is de radiodetectietechniek een uitstekend hulpmiddel voor calorimetrische metingen van de elektromagnetische componenten van de regen aan secondaire deeltjes veroorzaakt door kosmische stralen. Men kan het functioneren van de deeltjes detectoren controleren, omdat het

afgegeven signaal van deze detectoren afhangt van veroudering van de fotomultiplicator buis, de waterkwaliteit en reflectiviteit van de binnenzijde van deze detectoren. Hierdoor kunnen de deeltjes detectoren beter gecontroleerd worden door radio detectie technieken, vergeleken met fluorescentie detectie technieken, die afhankelijk zijn van atmosferische correcties en laserkalibraties. Om deze reden is AERAlet niet alleen een natuurkundig meetinstrument (uitbreiding naar lagere energieën), maar is de combinatie van AERAlet met AREA een kalibratie-instrument voor de elektromagnetische component van de deeltjes regen.

De geregistreerde radiosignalen door de 24 stations van de eerste fase van AERA, zijn gebruikt om het laterale gedrag van de polarisatie van het elektrisch veld in kaart te brengen, dat veroorzaakt wordt door de deeltjes regen. Deze informatie is direct gekoppeld aan het type van het emissie mechanisme. In onze analyse wordt de bijdrage aan het totale signaal van het radiaal gepolariseerde elektrische veld ten opzichte van de dominante geomagnetische bijdrage berekend. We vonden dat deze fractie in principe nul is op de plaats waar de deeltjes intensiteit maximaal is op het aardoppervlak. Op deze plaats is de emissie puur geomagnetisch. De radiale bijdrage begint significant te worden bij een laterale afstand van 90-100 m vanaf dit centrum. Helaas kunnen we dit alleen kwalitatief beschrijven. De onzekerheid op de gereconstrueerde geometrische hoeken, die de waarnemerspositie ten opzichte van het centrum karakteriseren, hebben een sterke invloed op de onzekerheid waarmee deze fractie bepaald kan worden. Deze onzekerheden hebben voornamelijk betrekking op de nauwkeurigheid van de bepaling van het maximum van de deeltjes regen op het aardoppervlak. Weliswaar is deze bepaling verbeterd door de AERAlet reconstructie, maar het blijft een dominante factor in de berekening van de polarisatie informatie uit de geregistreerde radio signalen.

Deze laterale afhankelijkheid van de radiaal-gepolariseerde bijdrage bevestigt de recente resultaten van het LOFAR experiment en draagt bij aan ons begrip van de radio-emissie mechanismen in deeltjes regens.

"As you set out for Ithaka
hope the voyage is a long one,
full of adventure, full of discovery.
Laistrygonians and Cyclops,
angry Poseidon – don't be afraid of them:
you'll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.
Laistrygonians and Cyclops,
wild Poseidon – you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.
Hope the voyage is a long one.
May there be many a summer morning when,
with what pleasure, what joy,
you come into harbors seen for the first time;
may you stop at Phoenician trading stations
to buy fine things,
mother of pearl and coral, amber and ebony,
sensual perfume of every kind – as many sensual perfumes as you can;
and may you visit many Egyptian cities
to gather stores of knowledge from their scholars.

Keep Ithaka always in your mind.
Arriving there is what you are destined for.
But do not hurry the journey at all.
Better if it lasts for years,
so you are old by the time you reach the island,
wealthy with all you have gained on the way,
not expecting Ithaka to make you rich.
Ithaka gave you the marvellous journey.
Without her you would not have set out.
She has nothing left to give you now.
And if you find her poor, Ithaka won't have fooled you.
Wise as you will have become, so full of experience,
you will have understood by then what these Ithakas mean."

(C.P. Cavafy, Collected Poems)